Corinna Forner

PRÜFUNG UND PRÜFBARKEIT VON HYPOTHESEN ÜBER MNESTISCH STABILISIERTE BEZUGSSYSTEME

1974

PSYCHOLOGIE

PRÜFUNG UND PRÜFBARKEIT VON HYPOTHESEN ÜBER MNESTISCH STABILISIERTE BEZUGSSYSTEME

Inaugural - Dissertation
zur Erlangung des Doktorgrades
der
Philosophischen Fakultät
der
Westfälischen Wilhelms-Universität zu
Münster (Westf.)

vorgelegt von

CORINNA FORNER, geb. ZAHN

aus HERFORD

1974

Tag der Promotion: 19. Juli 1974

Tag der mündlichen Prüfung: 16. Juli 1974

Dekan: Prof. Budde

Referent: Prof. Witte

Korreferent: Dr. Kallina

UNIV. LEIDEN BIBL.

Einleitung

Die Bezugssysteme und ihre Erforschung

I. Bezugssysteme 1

II. Bezugssystemforschung 4

Erster Teil:

Hypothesen, Methoden, Material

A. Die Aufgabenstellung dieser Arbeit 8

B. Die Hypothesen

I. Allgemeine Erläuterung der Begriffe 9

II. Erläuterung der Hypothesen 14

III. Die mathematischen Methoden für die Bearbeitung der Hypothesen 15

IV. Das Material, an dem die Überprüfung der Hypothesen versucht werden soll 16

C. Das Material

I. Methodenbedingte Gesichtspunkte für die Auswahl 17

II. Sichtung des vorhandenen Materials 19

Zweiter Teil:

Mathematisch-statistische Überprüfung

A. Prüfung auf Normalverteilung

I. Die statistische Bearbeitung

1. Einleitende Überlegungen 2o

2. Versuche mit Systemenden 33

a. Prüfung auf Normalverteilung bei Systemanfang und Systemende 39

b. Prüfung auf Normalverteilung beim absoluten Systemumfang 41

c. Prüfung auf Normalverteilung bei den relativen Bereichsumfängen 43

3. Versuche ohne Systemenden 44

II. Diskussion der Ergebnisse 45

1. Unterschied zwischen Systemanfang und -ende 46

2. Absoluter Systemumfang bei Versuchen mit einem und beiden Systemenden 49

III. Zusammenfassung der Ergebnisse und Folgerungen für spätere Untersuchungen 51

B. Prüfung auf Unabhängigkeit

I. Die statistische Bearbeitung 55

1. Einleitende Überlegungen 55

2. Prüfung auf Unabhängigkeit der relativen Bereichsumfänge vom absoluten Systemumfang 63

3. Prüfung auf Unabhängigkeit des Systemanfanges vom Systemende 67

4. Prüfung auf Unabhängigkeit der Systemenden von der relativen Bereichsgröße 7o

II. Diskussion der Ergebnisse 72

1. Korrelationen mit einer Konstanten 73

2. Zusammenhang anzeigende Koeffizienten 74

a. Korrelation zwischen relativer Bereichsgröße und Systemumfang 74

b. Korrelation zwischen Systemanfang und -ende 8o

c. Korrelation zwischen Systemenden und relativer Bereichsgröße 82

3. Mittelwert und Streuung 84

III. Zusammenfassung der Ergebnisse und Folgerungen für spätere Untersuchungen 86

C. Der Streuungsvergleich 87

I. Die statistische Bearbeitung und Diskussion der Befunde 89

1. Einleitende Überlegungen 91

2. Die Methode von BRÄUER 92

a. BRÄUER: Die Entwicklung von Bezugssystemen von der späten Kindheit an über Vorpubertät und Pubertät zum Erwachsenenalter 92

b. BRÄUERs Vorgehensweise 94

3. Eine Modifizierung der BRÄUERschen Methode 1oo

a. Die Versuche von BRÄUER 1oo

b. Die Versuche von HRUSCHKA 1o3

c. Unterschiede entsprechend der Kategorienzahl 1o8

4. Die Methode von Mittelwert und Streuung 112

II. Methodenvergleich, Zusammenfassung der Ergebnisse und Ausblick auf spätere Untersuchungen

1. Zusammenfassung 119

2. Vergleich der beiden Methoden 121

3. Ausblick auf spätere Untersuchungen 124

Dritter Teil:

Zusammenfassung 126

Literaturverzeichnis 136

Einleitung

Die Bezugssysteme und ihre Erforschung

I. Bezugssysteme

"Die Bezugssysteme sind die im Allgemeinen selbst nicht wahrgenommenen Grundlagen, durch die jedes psychische Gebilde (Wahrnehmung, Erlebnis, jede erlebbare Eigenschaft) seine spezifische Richtung, Größe, Stärke, Qualität usw. erhält." [1)]

Die Bezugssysteme vermitteln dem Menschen das Wissen um den Stellenwert eines Ortes und ermöglichen ihm so die Orientierung. Bezugssysteme existieren auf allen psychischen Gebieten, sie dienen der Orientierung sowohl auf räumlicher als auch auf geistiger Ebene. Ein Bezugssystem könnte als eine Art Meßstab angesehen werden, der aus der Fülle unserer bisherigen Erfahrung entstanden ist und den wir an unsere augenblicklichen Wahrnehmungen anlegen, um sie so in das Gesamt unserer psychischen Welt einzuordnen.

So haben wir im Laufe unseres Lebens eine Vielzahl von Bäumen gesehen, die von unterschiedlicher Größe waren. Alle diese uns bekannten möglichen Größenunterschiede bei den Bäumen haben wir in dem zugehörigen Bezugssystem vereinigt. Sehen wir nun einen Baum, so können wir sofort sagen, ob dieser Baum groß oder klein ist. Für dieses Urteil müssen wir nicht in Gedanken alle jemals gesehenen Bäume als Vergleich heranziehen. Auch beschränken wir uns bei diesem Urteil nicht auf die gerade sichtbaren Bäume und sagen, daß dieser Baum im Vergleich zu den anderen groß sei, sondern wir stellen ganz absolut fest, dieser Baum ist groß.

Die Erfahrung ist ein wichtiger Bestandteil unserer Bezugssysteme.

Wachsen wir in einer Gegend auf, in der es nur kleine, verkrüppelte Bäume gibt, so ist für uns ein Baum schon dann groß, wenn er für einen anderen Menschen, der bereits viel größere Bäume gesehen hat, noch klein ist. Außerdem spielt sie eine große Rolle für die Sicherheit, mit der wir uns unserer Bezugssysteme bedienen. Haben wir mit einer Sache erst wenige Erfahrungen gemacht, so können wir sie nur sehr unsicher beurteilen. Essen wir z.B. fremdländische Gerichte, so schmeckt alles erst einmal sonderbar und nur mit zunehmender Gewöhnung können wir ein Urteil über die Geschmackslage abgeben. Haben wir noch nie einen Baum gesehen, so können wir nicht beurteilen, ob der gerade gesehene groß oder klein ist. Wir können dann andere Bezugssysteme heranziehen und auf ihrer Grundlage zu einem Urteil gelangen. Kennen wir zwar Sträucher, aber keine Bäume, so sagen wir vielleicht, das ist aber ein großer Strauch, der etwas ungewöhnlich aussieht.

Wir besitzen für alles, was wir bislang kennengelernt haben, mehr oder minder fest ausgebildete Bezugssysteme. Diese Systeme bestehen nicht nebeneinander, sondern bilden ein Geflecht ineinandergreifender Eigenschaften. So können wir ein Urteil aufgrund verschiedener Bezugssysteme abgeben, die die einzelnen Aspekte des zu beurteilenden Gebildes berücksichtigen.

Begegnen wir einem Gebilde, das nicht in den Rahmen unserer Bezugssysteme paßt, so können wir kein spontanes Urteil abgeben. Kennen wir keine Bäume und sehen einen solchen zum ersten Mal, so können wir nicht sagen, ob es sich um einen großen oder kleinen handelt. Wenn wir jedoch sagen, das ist eine schöne oder eine ungewöhnliche Pflanze, so urteilen wir auf der Grundlage eines anderen Bezugssystems mit anderen absoluten Eigenschaften. Lernen wir nun immer mehr Bäume kennen, so entsteht auch das zugehörige Bezugssystem für die Größe von Bäumen. Kannten wir nur kleine Bäume und lernen dann größere kennen, so ändert sich unser Bezugssystem, indem es die gemachten Erfahrungen mit größeren Bäumen integriert.

"Ein seelisches Bezugssystem ist eine lebendige Ganzheit, die als solche auf jede auch nur örtliche Beanspruchung reagiert, und indem sie diese aufnimmt und bestimmt, umgekehrt auch von ihr beeinflußt und bestimmt wird." [2)]

II. Bezugssystemforschung

In das Blickfeld des Psychologen traten die Bezugssysteme durch eine Beobachtung von LILIAN MARTIN, einer Schülerin G.E. MÜLLERs. Bislang war man der Ansicht, daß es nur vergleichende Urteile gäbe, und L. MARTIN wollte diesen Vergleichsprozeß in ihren Versuchen näher untersuchen. Sie ließ ihre Versuchspersonen Gewichte vergleichen, wobei festgestellt werden sollte, ob das gerade gehobene Gewicht leichter oder schwerer als die bereits gehobenen Gewichte sei. Die so abgegebenen Urteile waren jedoch nicht alle 'komparativisch-relativ', d.h. vergleichender Natur, sondern zum Teil absolut. Mit zunehmender Versuchsübung urteilten die Versuchspersonen nämlich nicht immer mit leichter oder schwerer, sondern einfach mit leicht oder schwer. Sie machten beziehungslose, absolute Angaben.

Die Versuchspersonen haben die Erfahrungen mit den im Versuch verwendeten Gewichten zusammengefaßt, geordnet und zueinander in Beziehung gesetzt, es entstand ein Bezugssystem für die Schwere der Versuchsgewichte. Auf der Grundlage dieses entstandenen Bezugssystems wurden dann im Folgenden die Urteile abgegeben, deren Eigenschaften absolut waren.

Dieser Sachverhalt wiedersprach den damaligen Theorien über den Urteilsprozeß, die nur den vergleichenden Aspekt beim Zustandekommen eines Urteils berücksichtigten. Die Systeme, die den spontanen, nicht vergleichenden Urteilen zugrundelagen, wurden daraufhin als Bezugssysteme bezeichnet und deren Eigenschaften als absolut. Dies war der Anfang der Erforschung der Bezugssysteme.

Es galt nun festzustellen, ob Bezugssysteme die Ausnahme oder die Regel waren. Waren die absoluten Urteile bei den Gewichtsversuchen ein spezifisches Begleitmoment dieses Versuches oder eine allgemein bestehende Tatsache?

Als einmal der Anstoß gegeben war, entdeckte man, daß im täglichen Leben viel mehr absolut als vergleichend geurteilt wurde. An der Existenz von Bezugssystemen konnte bald nicht mehr gezweifelt werden. Man untersuchte im Folgenden einmal Bezugssysteme, die erst während des eigentlichen Versuchs entstanden, wie das Bezugssystem für die Schwere jener Gewichte, und zum anderen Bezugssysteme, die bereits vor den Versuchen ausgebildet waren. Erstaunlicherweise waren bei diesen vielfältigen Untersuchungen stets alle Versuchspersonen in der Lage, absolut zu urteilen. Es gab keine Schwierigkeiten, die Versuchsanweisungen zu befolgen, und die Versuchspersonen empfanden die geforderte Beurteilungsweise nicht als etwas Unnatürliches, mit dem man sie vielleicht überforderte. Die Bezugssysteme waren ein bislang übersehener, allgemein gebräuchlicher Bestandteil des menschlichen Urteils- und Orientierungsvermögens.

W. WITTE und seine Schüler nahmen sich der Problematik der Bezugssysteme eingehend an. Von Interesse war nun nicht mehr, die Existenz von Bezugssystemen nachzuweisen, sondern gewisse Eigenschaften dieser Systeme zu erkennen. Dafür war es notwendig, die Systeme zu untersuchen, deren man sich mit großer Sicherheit bediente, da ihre Eigenschaften stabil und gefestigt waren. Bezugssysteme, die während des Versuchs selbst entstehen, sind relativ flüchtig, jeder neue, dargebotene Reiz kann sie verändern. Es mußten Bezugssysteme untersucht werden, die schon vor dem Versuch bestehen und gefestigt, und so durch den Versuch selbst nicht mehr zu beeinflussen sind. Hat man die Eigenschaften dieser Systeme erforscht und kennt man ihre Gesetzmäßigkeiten, so kann man sich den weniger stabilen Systemen zuwenden und auf der Grundlage der gemachten Erkenntnisse die Eigenschaften dieser Systeme in Abhebung zu den anderen feststellen.

Man beschäftigte sich so im Folgenden mit den mnestisch stabilisierten Bezugssystemen, das sind die Systeme, die durch Erfahrung und ständige

Übung ein fester und stabiler Bestandteil des Menschen geworden sind. Besonders häufig wurde damals das Bezugssystem der Bleistiftgröße untersucht. Bleistifte eigneten sich sehr gut, da sie einmal wenig Versuchsaufwand benötigten, zum anderen wohl allen Versuchspersonen geläufig waren.

Man versuchte dann, ein möglichst breit gestreutes Feld zu untersuchen. Die Bleistiftgrößen variierten im eindimensionalen Bereich. Andere Versuche beschäftigten sich mit Gegebenheiten, die in mehrdimensionalen Bereichen variierten, so für den zweidimensionalen Bereich Untersuchungen über die Größe von Taschentüchern (Quadrate) und Armbanduhren (Kreise), im dreidimensionalen Bereich über die Größe von Christbaumkugeln (Kugel).

Außer dem visuellen Bereich wurde auch der haptische und akustische Bereich herangezogen, wie bei Untersuchungen über die Schwere von Äpfeln oder die Tonhöhen einzelner Töne verschiedenster Instrumente.

Allen Versuchen gemeinsam war, daß sie sich mit mnestisch stabilisierten Systemen befaßten. Variiert wurde die Gegebenheit, die zu beurteilen war und die Anzahl oder Art der Kategorien, die für die Beurteilung erstellt wurden. Ziel dieser Untersuchungen war es, einen möglichst weitgefächerten Katalog untersuchter Bezugssysteme anzulegen, um auf vielen Gebieten festzustellen, wie die Versuchspersonen zu ihren absoluten Urteilen gelangen und was sie eventuell beeinflussen könnte. Aufgabe der Bezugssystemforschung ist es nun, die vielen Einzelergebnisse zusammenzufassen und allgemeingültige Gesetzmäßigkeiten zu formulieren, die dann Gegenstand zukünftiger, gezielter Untersuchungen sein werden.

WITTE hat von Anfang an die Entwicklung der Forschung gelenkt und verfolgt und hat auf der Grundlage dieses Wissens einige allgemeingültige Gesetzmäßigkeiten zu formulieren gesucht. Die Überprüfung

dieser Gesetzmäßigkeiten steht nun im Mittelpunkt der Bezugssystem-forschung und diese Arbeit soll einen Beitrag dazu leisten.

Erster Teil:

Hypothesen, Methoden und Material

A. Die Aufgabenstellung dieser Arbeit

WITTE formulierte auf der Grundlage der bisher gemachten Untersuchungen und deren Erkenntnisse folgende allgemeingültige Gesetzmäßigkeiten [3] :

1. Die Systeme sind ihrem absoluten Umfang ebenso wie die Bereiche ihrem relativen Umfang nach normalverteilt.

2. Die Systemenden sind normalverteilt.

3. Jeder relative Bereichsumfang ist unabhängig vom absoluten Systemumfang.

4. Systemanfang und -ende sind unabhängig voneinander.

5. Systemanfang und -ende sind unabhängig von den relativen Bereichsgrößen.

6. Die intraindividuelle Streuung ist überaus signifikant kleiner als die interindividuelle Streuung.

Die Gültigkeit dieser Gesetzmäßigkeiten ist bislang noch nicht überprüft worden, es handelt sich deshalb hier um Hypothesen, deren Beantwortung noch aussteht.

Aufgabe dieser Arbeit ist es nicht, eine Hypothese herauszugreifen und anhand dafür unternommener Versuche zu verifizieren oder falsifizieren. Bei dieser Arbeit steht das Gesamt der Hypothesen zur Diskussion. An den bereits gemachten Untersuchungen und deren Be-

funde soll eine mathematisch-statistische Überprüfung versucht werden. Wo das Material für eine solche Überprüfung nicht ausreicht, werden Überlegungen anzustellen sein, welche Vorraussetzungen für eine den Anforderungen entsprechende Behandlung geschaffen werden müssen.

Die Arbeit schließt an die bisherigen Untersuchungen an, indem sie deren Befunde zum Gegenstand der statistischen Überprüfung macht, sie ist aber bereits Teil der neuen Phase der Bezugssystemforschung, in deren Mittelpunkt die gezielte Untersuchung von Hypothesen steht. Die Hypothesen werden anhand des vorhandenen Materials nicht erschöpfend zu behandeln sein. Zentrales Thema dieser Arbeit ist es deshalb auch, die Gesichtspunkte herauszuarbeiten, die von den Hypothesen her bedingt sind, und die Versuchsbedingungen zu untersuchen, die eine Behandlung der Hypothesen beeinflussen können.

Der Aufbau der Arbeit wird von den zur Diskussion stehenden Hypothesen und den statistischen Verfahren zu deren Beantwortung bestimmt sein.

B. Die Hypothesen

I. Allgemeine Erläuterung der Begriffe

Bei den von WITTE formulierten Gesetzmäßigkeiten wird eine Reihe von Begriffen verwendet, deren Verständnis für den weiteren Verlauf dieser Untersuchung notwendig ist.

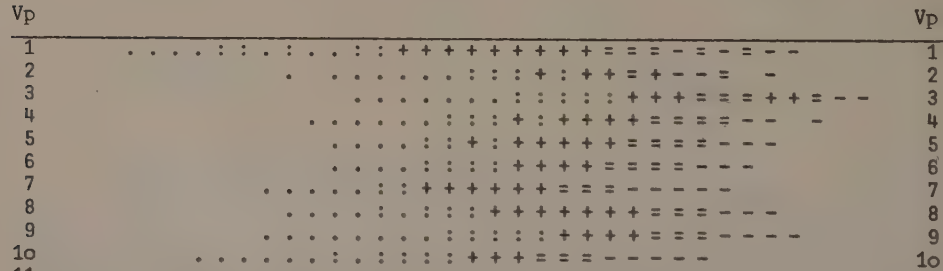
Folgende Begriffe bedürfen einer näheren Erklärung:
System, absoluter (System)Umfang, Bereiche, relativer Bereichsumfang, Systemenden, Systemanfang und -ende, intra- und interindividuelle Streuung.

Diese Begriffe kann man am besten an einem Versuch selbst veranschaulichen. Wir wollen dafür eine Untersuchung von HRUSCHKA [4] nehmen:

Es handelt sich um einen Versuch über die Größenbeurteilung von Armbanduhren. Als Versuchsvorlage dienten 38 Fotographien von runden Armbanduhren, deren Durchmesser äqidistant um einen cm zunahm. Der Versuch wurde mit 15 männlichen Versuchspersonen (Vpn) durchgeführt. Für die Beurteilung standen folgende fünf Kategorien zur Verfügung: sehr groß, groß, mittel, klein und sehr klein. Die Versuchsvorlagen sollten Herrenuhren darstellen und dementsprechend beurteilt werden. Vorlagen, die in den Rahmen von Herrenuhren nicht paßten, weil sie entweder zu groß oder zu klein waren, sollten nicht beurteilt werden. Die einzelnen Versuchsvorlagen wurden im Konstanzverfahren dargeboten.

Die Ergebnisse dieses Versuchs sind in der folgenden Graphik zusammengefaßt:

Bezugssystem : Armbanduhren (Herren), Einteilung in fünf Bereiche



Die Graphik stellt eine Matrix dar, in der die Urteile aller Vpn zu jeder Versuchsvorlage wiedergegeben sind. Auf der Abszisse sind die Durchmesser der Uhren, von 5o bis 13 cm variierend, aufgetragen, auf der Ordinate die Nummern der einzelnen Vpn von 1 bis 15.

Die einzelnen Kästchen geben das Urteil der entsprechenden Vp zu der entsprechenden Versuchsvorlage wieder. Die bei der Beurteilung gewählten Kategorien werden durch die unterschiedlichen Zeichen (siehe Zeichenerklärung) in den einzelnen Kästchen dargestellt. Kästchen, in denen kein Zeichen ist, bedeuten, daß diese Versuchsvorlage von der jeweiligen Vp als nicht mehr zur Mannigfaltigkeit der 'Herrenuhren' gehörend angesehen wurde.

Das System oder Bezugssystem einer Vp erstreckt sich vom größten bis zum kleinsten Urteil. Die Versuchsvorlagen, die nicht mit einem kategorialen Urteil bedacht wurden, gehören nicht zum System dieser Vp. Jede Vp hat ihr persönliches Bezugssystem. Was eine Vp noch als große Herrenuhr ansieht, liegt für eine andere Vp schon längst nicht mehr in ihrem System 'Herrenuhr'.

Am Anfang oder Ende eines Systems treten manchmal Lücken auf, das System erstreckt sich nicht kontinuierlich vom größten bis zum kleinsten akzeptierten Reiz; diese Lücken treten ebenfalls bei der Einteilung in verschiedene Kategorien auf. Dies sind Zeichen für Unsicherheit bei der Beurteilung, die wohl z.T. durch die Konstanzmethode bedingt sind. Die Vpn kennen vor Beginn des Versuchs die Variationsbreite der Versuchsvorlagen nicht. Das Auftreten von z.B. noch kleineren Uhren im späteren Verlauf des Versuchs kann dazu verleiten, daß eine Uhr noch akzeptiert wird, obwohl die nächst größere bereits vorher abgelehnt wurde. Bei Versuchswiederholungen wird sich die Beurteilung stabilisieren, da dann keine unbekannten Größen mehr vorhanden sind. Dieses Vorhandensein von Lücken zeigt aber auch, daß die Systeme der einzelnen Vpn nicht so sehr stabilisiert sind, daß der Versuch selbst sie nicht mehr beeinflussen könnte.

Für die hier zur Frage stehenden Untersuchungen ist jedoch die absolute Lage der Systeme nicht von Wichtigkeit, es spielt keine Rolle, ob das System nun bei der Vorlage mit 42 cm Durchmesser oder 41 cm Durchmesser beginnt. Wichtig ist hier die Größe der einzelnen Systeme, d.h. wieviele Versuchsvorlagen individuell mit kategorialen Urteilen bedacht wurden. Das Problem der Lücken und Unregelmäßigkeiten kann deshalb außer Acht gelassen werden.

Mit absolutem Systemumfang bezeichnet man die Anzahl der kategorial abgegebenen Urteile oder die Anzahl der Versuchsvorlagen, die zu dem jeweiligen System einer Vp gehören (was das Gleiche bedeutet). Vp 1 in unserem Beispiel hat 3o kategoriale Urteile abgegeben, ihr absoluter Systemumfang beträgt 3o.

Innerhalb eines Systems unterscheidet man einzelne Bereiche, die von den zur Verfügung stehenden Kategorien eingenommen werden. Jede Kategorie wurde bei der Beurteilung mit einer bestimmten Häufigkeit gewählt. So hat Vp 1 7mal mit 'sehr groß', 5mal mit 'groß', 9mal mit 'mittel', 5mal mit 'klein' und 4mal mit 'sehr klein' geurteilt. Die Größe der einzelnen Bereiche entspricht der Anzahl der Urteile, die in der zugehörigen Kategorie abgegeben wurden. Der Bereich der Kategorie 'sehr groß' erstreckt sich bei der Vp 1 über 7 Urteile, die Bereichsgröße beträgt 7. Da die Systeme der einzelnen Vpn von unterschiedlicher Größe sind, sagt die absolute Größe eines Bereiches nichts darüber aus, ob dieser Bereich besonders groß oder klein innerhalb des Systems ausfällt.

Vp 1 und Vp 2 haben jeweils 7 sehr-groß-Urteile abgegeben, der absolute Systemumfang der ersten Vp beträgt 3o, der der zweiten jedoch nur 2o. 7 von 3o oder 7 von 2o Urteilen bedeutet aber einen unterschiedlich großen Bereich innerhalb des Systems der beiden Vpn. Bei der ersten Vp umfaßt der Bereich dieser Kategorie 23,3%, bei der zweiten Vp 35% der insgesamt abgegebenen Urteile. Um die Bereichs-

größen der einzelnen Vpn miteinander vergleichen zu können, ist es notwendig, diese in Beziehung zum zugehörigen absoluten Systemumfang zu setzen, womit man den relativen Bereichsumfang erhält. Diese Prozedur ist deshalb notwendig, weil man unter Bereich nicht die Anzahl der abgegebenen Urteile versteht, sondern das Ausmaß, das diese Kategorie innerhalb des Systems einnimmt.

Bei Systemen, die sich nicht über die ganze Skala der dargebotenen Reize erstrecken, gibt es jeweils einen Anfang und ein Ende des Systems. Ob man nun den Groß- oder den Kleinbereich als Anfang bezeichnet, ist gleichgültig. Wichtig ist jedoch, daß man den einmal gewählten Anfangspunkt bei allen Untersuchungen beibehält, da sonst mit einem Namen unterschiedliche Tatbestände bezeichnet werden.

Um ein von der Lage des Systems selbst unabhängiges Maß für die Systemenden zu erhalten, zählt man für jede Vp aus, wie viele Versuchsvorlagen sie rechts und links von ihrem System nicht mehr akzeptierte. Da allen Vpn gleich viele Gegebenheiten zur Beurteilung vorgelegt worden sind, haben die absoluten Werte dieselbe Aussagekraft wie auf die Gesamtskala bezogene Werte (die Gesamtskala ist ein nicht variierender konstanter Wert). Legt man den Anfang eines Systems im Großbereich der dargebotenen Vorlagen und das Ende im Kleinbereich fest, so bedeutet Systemanfang die Anzahl der Vorlagen, die für das zugehörige System als zuklein erachtet wurden, Systemende die Anzahl der Vorlagen, die als zu groß erachtet wurden. Bei Vp 1 beträgt demnach der Systemanfang 5, das Systemende 3. Es bestehen nun folgende Beziehungen zwischen den einzelnen Größen:

Die Summe der absoluten Bereichsumfänge ist gleich dem absoluten Systemumfang; die Summe der relativen Bereichsumfänge ist gleich 100%; die Summe von absolutem Systemumfang, Systemanfang und Systemende ist gleich der Anzahl der insgesamt dargebotenen Versuchsvorlagen oder der Variationsbreite der Vorlagen.

Die Beurteilungen einzelner Vpn unterscheiden sich voneinander. Nimmt man als Bezugspunkt eine bestimmte Versuchsvorlage, so wird diese von den Vpn z.T. unterschiedlich eingestuft. Die Uhr, die einen Durchmesser von 32 cm hat, wird in unserem Versuch von einigen Vpn mit 'mittel', von anderen mit 'groß' und von noch anderen mit 'sehr groß' bezeichnet. Die Unterschiede von Vp zu Vp werden interindividuelle Streuung genannt. Die Unterschiede bei der Beurteilung ein und derselben Versuchsvorlage von einer Vp werden als intraindividuelle Streuung bezeichnet. Bei unserem Versuch gibt es keine intraindividuelle Streuung, da jede Vp nur ein Urteil zu jeder Vorlage abgegeben hat. Wiederholt man den Versuch unter gleichen Bedingungen, so erhält man für jede Vp jeweils zwei Beurteilungen zu einer Vorlage, man kann eine intraindividuelle Streuung feststellen.

II. Erläuterung der Hypothesen

1. Die Systeme sind ihrem absoluten Umfang ebenso wie die Bereiche ihrem relativen Umfang nach normalverteilt.

   Diese Hypothese besagt, daß die absoluten Umfänge der Systeme, die von Vp zu Vp unterschiedlich sind, ihrer Verteilung nach einer Normalverteilung entsprechen. Die einzelnen Bereiche innerhalb der Systeme sollen sich ebenfalls normalverteilen, d.h. der relative Bereichsumfang für die erste, zweite oder dritte Kategorie bei allen Vpn bildet eine Normalverteilung.

2. Die Systemenden sind normalverteilt.

   Die Verteilung der Systemanfänge und die Verteilung der Systemenden soll ebenfalls einer Normalverteilung entsprechen.

3. Jeder relative Bereichsumfang ist unabhängig vom absoluten Systemumfang.

   Dieser Satz besagt, daß die Größe eines Bereiches nicht von der Größe des gesamten Systems abhängt, daß z.B. ein, im Vergleich zu

anderen Vpn gesehen, kleines System nicht bedingt, daß die relativen Bereichsgrößen innerhalb des Systems ebenfalls klein sein müssen.

4. Systemanfang und -ende sind unabhängig voneinander.

   Die Anzahl der Gegebenheiten, die vor dem Beginn eines Systems liegen, hat keinen Einfluß auf die Anzahl der Gegebenheiten, die hinter einem System liegen.

5. Systemanfang und -ende sind unabhängig von den relativen Bereichsgrößen.

   Die Anzahl der Gegebenheiten, die vor oder hinter einem System liegen, haben keinen Einfluß auf die relative Größe der einzelnen Bereiche.

6. Die intraindividuelle Streuung ist überaus signifikant kleiner als die interindividuelle Streuung.

   Bei Versuchen, bei denen einzelne Versuchsdurchgänge wiederholt wurden, streuen die Urteile einer Vp von Wiederholung zu Wiederholung weitaus weniger als die Urteile der Vpn unter sich.

III. Die mathematischen Methoden für die Bearbeitung der Hypothesen.

Die ersten beiden Hypothesen sagen Normalverteilung für bestimmte Verteilungen (der relativen Bereichsumfänge, der Systemenden) voraus. Für die in einem Versuch gefundene Verteilung kann eine Normalverteilung mit den selben Parametern berechnet werden. Zu prüfen ist dann, wie gut die gefundene Verteilung der berechneten Normalverteilung entspricht, in wie weit beide Kurven einander gleichen. Für den Verteilungsvergleich steht uns als Methode das Chi-Quadrat zur Verfügung.

Die dritte, vierte und fünfte Hypothese postuliert Unabhängigkeit zwischen zwei Größen (relativer Bereichsumfang - absoluter Systemumfang, Systemanfang - Systemende, Systemanfang - relativer Bereichs-

umfang, Systemende - relativer Bereichsumfang). Ein Maß für Abhängigkeit zwischen zwei Größen ist der Korrelationskoeffizient. Liegt dieser um Null herum, so besagt er, keine Abhängigkeit zwischen diesen Größen. Um Unabhängigkeit feststellen zu können, muß also auf Nullkorrelation hin geprüft werden.

Die sechste Hypothese macht eine Aussage über zwei Streuungen. Die erste Aufgabe ist es hier, ein adäquates Maß für die einzelnen Streuungen zu finden. Danach kann man dann die intraindividuellen Streuungen mit den interindividuellen vergleichen und mit einem Paardifferenzverfahren feststellen, ob sich beide Wertegruppen voneinander unterscheiden.

IV. Das Material, an dem eine Überprüfung der Hypothesen versucht werden soll.

Die oben genannten statistischen Verfahren sollen auf die Befunde bereits durchgeführter Versuche angewendet werden. Dafür müssen die Untersuchungen herausgegriffen werden, bei denen das erzielte Datenmaterial eine solche Überprüfung zuläßt. Es muß dabei stets berücksichtigt werden, daß die Versuche, deren Befunde herangezogen werden, nicht für die Behandlung dieser Hypothesen angelegt wurden, sondern umgekehrt die Hypothesen aus den Befunden dieser Untersuchungen hervorgingen. Kritische Äußerungen über die Brauchbarkeit der Befunde oder über die Adäquatheit der Versuchssituation und der Durchführung sollen deswegen nicht zu Lasten der Versuche gehen, sondern nur Anhaltspunkte für spätere Untersuchungen sein.

Die Spezifität der Hypothesen stellt gewisse Grundvoraussetzungen an das zu verwendende Datenmaterial und die Auswahl wurde nach diesen Gesichtspunkten getroffen. Eine Wertung der Versuche ist damit nicht gemeint.

Einige grundlegende mathematische Fragen bezüglich des Datenmaterials, wie z.B. die Größe der Kategorialabschnitte oder die Lage der Häufungspunkte in den Bereichen, stehen hier nicht zur Diskussion und werden im weiteren Verlauf nicht behandelt. Der daran interessierte Leser möge in den Orginalarbeiten nachsehen, in denen diese Punkte abgehandelt werden.

C. Das Material

I. Methoden- und hypothesenbedingte Gesichtspunkte bei der Auswahl des Materials

Aus der Vielzahl der durchgeführten Versuche zur Bezugssystemproblematik müssen die Untersuchungen herausgesucht werden, die sich für die hier anstehende Überprüfung eignen. Die Anwendung der Prüfung auf Normalverteilung und Nullkorrelation und des Streuungsvergleiches bedingt numerische Daten als Grundlage. Diese Bedingung ist bei allen Bezugssystemversuchen, bei denen kategoriale Urteile zu verschiedenen Gegebenheiten abgegeben wurden, erfüllt, da man die Werte für die einzelnen Größen durch die Auszählung der Urteile erhält. Methodische Einschränkungen sind demnach vorerst nicht zu machen.

Durch die Spezifität der Hypothesen sind jedoch von vorneherein gewisse Einschränkungen zu machen. Die Hypothesen machen Aussagen über ganz spezielle Größen, die nicht bei allen Versuchen ermittelt werden können. Ganz generell kann man bei den bisher gemachten Untersuchungen zwischen zwei Arten unterscheiden: einmal Versuche mit Systemenden und zum anderen, in der Mehrzahl der Fälle, Versuche ohne Systemenden.

Für Versuche mit Systemenden haben wir bereits ein Beispiel ausführlich behandelt, nämlich den Versuch von HRUSCHKA mit den Herrenarmbanduhren. Wäre nun bei diesem Versuch die Variationsbreite der einzelnen Versuchsvorlagen derart eingeschränkt worden, daß keine Uhr bei keiner Vp außerhalb des Systems gelegen hätte, es also nur zu kategorialen Beurteilungen gekommen wäre, so hätte es sich um einen Versuch ohne Systemenden gehandelt. Bei diesen Versuchen kann man weder den absoluten Systemumfang, die eigentlichen Systeme erstrecken sich ja teilweise weit über den vorgegebenen Bereich hinaus, noch Systemanfang und -ende berechnen. Zu berechnen sind lediglich die Bereichsumfänge, wobei es hier keine Rolle spielt, ob man die absoluten oder die bezogenen Werte nimmt, da die Bezugsgröße für alle Vpn gleich ist, nämlich die Größe der vorgegebenen Variationsbreite der Vorlagen.

Mit dem Datenmaterial der Versuche ohne Systemenden kann nur ein Aspekt der ersten Hypothese behandelt werden, u. zw. die Normalverteilung der relativen Bereichsgrößen. Für die Überprüfung dieses Aspektes ist es jedoch wünschenswert, ebenfalls, wenn möglich, Daten von Versuchen mit Systemenden heranzuziehen, da die Bereichseinteilung durch die Einschränkung auf den für alle Vpn akzeptablen Mittelbereich ihrer Systeme beeinflußt sein könnte.

Die Prüfung auf Normalverteilung bei den absoluten Systemumfängen und den Systemenden und die Prüfung auf Nullkorrelation zwischen System- und Bereichsumfang, zwischen den Systemenden und zwischen Systemenden und Bereichsumfang kann nur anhand von Daten vorgenommen werden, die von Versuchen mit Systemenden stammen.

Bei der Hypothese über die Streuungen wird keine Aussage über bestimmte Größen, wie Systemumfang oder Systemenden gemacht. Von daher sind keine Einschränkungen zu machen. Die Voraussetzung für die Ermittlung der intraindividuellen Streuung ist jedoch in jedem Fall eine Versuchswiederholung, um wenigstens jeweils zwei Urteile pro Vp je Gegebenheit zu erhalten.

Zusammenfassend benötigen wir für die Bearbeitung der ersten fünf Hypothesen Daten von Versuchen mit Systemenden, ausgenommen für die Prüfung auf Normalverteilung bei den Bereichsumfängen. Für die Bearbeitung der sechsten Hypothese sind Daten von Versuchen erforderlich, bei denen wenigstens eine Versuchswiederholung durchgeführt wurde, u.zw. unter möglichst gleichbleibenden Bedingungen.

II. Sichtung des vorhandenen Materials

Durch die hypothesenbedingten Einschränkungen schrumpft das zur Verfügung stehende Material stark zusammen. Versuche mit Systemenden wurden nur von HRUSCHKA [4] durchgeführt. Versuche mit Versuchswiederholungen wurden zum Teil ebenfalls bei HRUSCHKA gemacht, durchgängig jedoch bei BRÄUER [5], hier aber ohne Systemenden. Für die Prüfung auf Normalverteilung bei den Bereichsumfängen wurden nicht alle Versuche, die bereits gemacht wurden, herangezogen, da dies einen zu großen Aufwand im Vergleich zum Stellenwert dieses Aspektes innerhalb der gesamten Überprüfung darstellen würde. Da auch hier besser mit den Daten von Versuchen mit Systemenden zu arbeiten ist, wurde dieser Aspekt einmal an den Daten von HRUSCHKA mit Systemenden bearbeitet, zum anderen an einer Auswahl von Versuchen ohne Systemenden. Folgende Versuche wurden dafür herangezogen: BUDDE [6], WINKELMANN [7], ALTRICHTER [8], STEINER [9].

Die einzelnen Untersuchungen berücksichtigen eine Vielzahl von Gegebenheiten, die im Folgenden aufgeschlüsselt worden ist:

HRUSCHKA	Bleistifte, Taschentücher, Armbanduhren, Christbaumkugeln
BRÄUER	Bleistifte, Klaviertöne
BUDDE	Waldhorn-, Pikkoloflöten-, große Flöten- und Posaunentöne
WINKELMANN	Bleistifte, Bauklötze
ALTRICHTER	arithmetisch abgestufte Grauscheiben
STEINER	geometrisch abgestufte Grauscheiben.

Zweiter Teil:

Mathematisch-statistische Überprüfung

A. Prüfung auf Normalverteilung

I. Die statistische Bearbeitung
1. Einleitende Überlegungen

Hypothese 1 und 2 postulieren Normalverteilung des absoluten Systemumfanges, der relativen Bereichsumfänge und der Systemenden.

Systemumfang und Systemenden erhält man nur bei Versuchen mit Systemenden. Diese Größen (und natürlich auch die relativen Bereichsumfänge) können nur an dem Datenmaterial von HRUSCHKA überprüft werden, da bei dieser Untersuchung mit Systemenden gearbeitet wurde (siehe S. 19). Die Normalverteilung der Bereichsumfänge kann außerdem noch an Versuchsdaten von Untersuchungen ohne Systemenden überprüft werden. Bei der statistischen Bearbeitung wird deshalb getrennt zwischen Versuchsdaten von Versuchen mit Systemenden (2.) und ohne Systemenden (3.).

Die Prüfung auf Normalverteilung kann mit Hilfe des Chi-Quardrats durchgeführt werden. Diese Methode vergleicht zwei Verteilungen hinsichtlich ihrer Ähnlichkeit, d.h. man kann mit ihr feststellen, ob beide der gleichen Grundgesamtheit entstammen und ihre Abweichungen nur zufällig sind oder ob beide Verteilungen verschiedenen Grundgesamtheiten zugehören.

Für die Anwendung des Chi-Quadrats benötigen wir zuerst zwei Verteilungen, die miteinander verglichen werden können. Unsere Daten liefern uns nur eine Verteilung, von der wir wissen wollen, ob sie einer Normalverteilung entspricht. Für unsere beobachtete Verteilung können wir den Mittelwert und die Streuung berechnen und auf der Grundlage dieser Kennwerte ist es möglich, die zugehörige Normalverteilung zu ermitteln. Diese berechnete Normalverteilung enthält die methodischen Eigenschaften

einer Normalverteilung ohne jegliche Zufallsschwankungen. Diese 'ideale' Verteilung kann nun mit unserer beobachteten Verteilung verglichen werden und durch das Chi-Quadrat erhalten wir dann einen Wahrscheinlichkeitswert über die zufälligen Schwankungen der beobachteten Verteilung. Das Chi-Quadrat sagt, mit wieviel prozentiger Wahrscheinlichkeit die Abweichungen der beobachteten Verteilung durch den Zufall bedingt sind, denn Abweichungen gibt es immer, da man in der Natur keine mathematisch exakten Verteilungen vorfinden kann.

Die beobachteten Verteilungen können wir nicht direkt von unserem Datenmaterial ablesen, sondern wir müssen dieses zuerst umwandeln. Die Rohdaten liegen uns in der Form vor, wie es im Beispiel auf Seite 1o zu sehen ist. Bei diesem Beispiel handelt es sich um einen Versuch mit Systemenden und wir können alle drei Größen erhalten, den absoluten Systemumfang, die relativen Bereichsumfänge und die Systemenden.

Als erstes müssen wir die Information der Matrix in numerische Werte umformen, d.h. wir müssen die Beurteilungen auszählen. Wir zählen so aus, wie oft jede Vp jede Kategorie gewählt hat und erhalten die absoluten Bereichsumfänge, deren Summe den absoluten Systemumfang ergibt. Ebenso zählen wir aus, wie oft die einzelnen Vpn vor ihrem System und hinter diesem Gegebenheiten als nicht zum System gehörend betrachteten, und erhalten so den Systemanfang und das Systemende. Für die Systemenden einigen wir uns darauf, daß der Systemanfang im Kleinbereich und das Systemende im Großbereich liegen sollen, d.h. Gegebenheiten, die für das eigentliche System zu klein sind, werden als Systemanfang bezeichnet, und Gegebenheiten, die für das System zu groß sind, als Systemende.

Wenn wir nun bei unserem Beispiel die Information umwandeln, d.h. mit der Auszählung beginnen, so erhalten wir folgende Tabelle:

Tabelle 1

Bezugssystem: Armbanduhren (Herren)
Auszählung der Urteile

Vp	S.A.	S.E.	a.S.U.	1.K.	2.K.	3.K.	4.K.	5.K.
1	5	3	3o	4	5	9	5	7
2	7	11	2o	3	2	4	4	7
3	2	13	23	2	4	5	5	7
4	5	11	22	3	4	5	4	6
5	6	12	2o	3	4	6	3	4
6	7	12	19	3	4	4	4	4
7	8	9	21	5	3	6	2	5
8	6	1o	22	3	3	7	4	5
9	5	9	24	4	3	4	5	8
1o	9	6	23	5	3	3	5	7
11	8	11	19	3	2	6	4	4
12	2	16	2o	3	3	5	6	3
13	3	11	24	1	6	8	6	3
14	8	12	18	1	5	5	4	3
15	7	14	17	2	4	4	5	2

Legende

S.A. : Systemanfang
S.E. : Systemende
a.S.U. : absoluter Systemumfang
1.K. : 1. Kategorie = sehr klein
2.K. : 2. Kategorie = klein
3.K. : 3. Kategorie = mittel
4.K. : 4. Kategorie = groß
5.K. : 5. Kategorie = sehr groß.

Die Werte der Systemenden und des absoluten Systemumfanges müssen nicht weiter umgewandelt werden, sie entsprechen den Werten unserer beobachteten Verteilung. Bei den Werten der Bereichsumfänge handelt es sich um die absoluten Werte, die noch auf den Systemumfang bezogen werden müssen, da in der Hypothese von den relativen Bereichsumfängen die Rede ist (siehe S. 12).

Die relativen Bereichsumfänge erhalten wir, indem wir die Größe der einzelnen Bereiche prozentual auf den zugehörigen absoluten Systemumfang beziehen. Die folgende Tabelle gibt die relativen Bereichsumfänge wieder:

Tabelle 2

Bezugssystem: Armbanduhren (Herren)
relative Bereichsumfänge

Vp.	1. Kat.	2. Kat.	3. Kat.	4. Kat.	5. Kat.
1	13	17	3o	17	23
2	15	1o	2o	2o	35
3	9	18	22	22	29
4	14	18	23	18	27
5	15	2o	3o	15	2o
6	16	21	21	21	21
7	24	14	28	1o	24
8	14	14	31	18	23
9	17	13	17	19	34
1o	22	13	13	22	3o
11	16	1o	32	21	21
12	15	15	25	3o	15
13	4	26	31	26	13
14	6	28	28	22	16
15	12	24	24	28	12

Legende:

1. Kat. : sehr klein
2. Kat. : klein
3. Kat. : mittel
4. Kat. : groß
5. Kat. : sehr groß.

Jetzt haben wir unsere Ausgangsdaten so weit umgeformt, daß wir mit der eigentlichen Berechnung beginnen können. Zuerst müssen wir die beobachteten Verteilungen ermitteln. Die Verteilung, die wir hier benötigen, ist eine Häufigkeitsverteilung. Wählen wir als Beispiel den relativen Bereichsumfang der ersten Kategorie, so müssen wir heraussuchen, wie häufig die einzelnen Prozentwerte vorkommen. Der niedrigste Prozentwert ist 4% und er kommt nur einmal vor, seine Häufigkeit ist also eins. Wir ordnen so der Größe nach die Prozentwerte und rechnen zusammen wie häufig dieser Wert vorkommt:

Tabelle 3

Bezugssystem: Armbanduhren (Herren)
Häufigkeitsverteilung des ersten relativen Bereichsumfanges

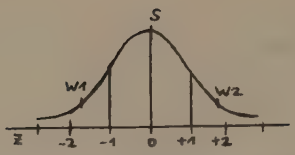
%	4	5	6	7	8	9	1o	11	12	13	14	15	16	17	18	19	2o	21	22	23	24
f	1	o	1	o	o	1	o	o	1	1	2	3	2	1	o	o	o	o	1	o	1

Die Häufigkeitsverteilung ist nun von 4% beginnend bis 24% endend nicht kontinuierlich besetzt, d.h. zwischen den Prozentwerten, die in unserer Tabelle 2 vorkommen, liegt eine Reihe von Prozentwerten, deren Häufigkeit gleich Null ist. Um die Chi-Quadrat-Prüfung auf Normalverteilung anwenden zu können, muß diese Verteilung in Klassen eingeteilt werden, damit eine kontinuierliche Skala entsteht.

Beginnen wir mit der Klasseneinteilung, so treten methodische Schwierigkeiten auf. Für die Anwendung des Chi-Quadrat-Tests wird gefordert, daß bei Klasseneinteilung die niedrigste Häufigkeit pro Klasse 5 betragen soll. Die Klassengröße muß selbstverständlich bei allen Klassen gleich sein. Versuchen wir bei unserer Häufigkeitsverteilung eine Klasseneinteilung mit minimalster Häufigkeit von 5, so können wir bei gleich großen Klassen lediglich in zwei Klassen einteilen, da sonst immer eine Klasse weniger als 5 Häufigkeiten in sich vereinigt. Mit 2 Klassen kann man aber nicht auf Normalverteilung prüfen. Die Normalverteilungskurve besitzt einen ganz spezifischen Verlauf, der in der Klasseneinteilung berücksichtigt werden muß:

Klasseneinteilung bei der Normalverteilung

Normalverteilungen haben folgende Eigenschaften:
Ihre Häufigkeiten verteilen sich gleichmäßig nach links und rechts um einen häufigsten Wert, der gleichzeitig Mittelwert und Median der Verteilung ist. Transformiert man die Maßzahlen X in Standardwerte, so haben alle Verteilungen den gleichen Verlauf der Häufigkeiten. Die Verteilung ist durch μ und σ völlig definiert. Die graphische Darstellung der Normalverteilung ist die Gauss-Kurve. Diese Kurve beginnt im negativen Unendlichen der X-Achse, verläuft praktisch asymptotisch zu dieser bis sie zum ersten Wendepunkt (W1) ansteigt. Am Wendepunkt geht die Linkskrümmung in eine Rechtskrümmung über und die Kurve steigt steil bis zum Sattelpunkt (S) an, der im Nullpunkt liegt. Vom Sattelpunkt fällt die Kurve steil zum zweiten Wendepunkt (W2) ab, wo die Rechtskrümmung nun in eine Linkskrümmung umwechselt. Danach fällt die Kurve langsam ab, sie wird praktisch wieder Asymptote der X-Achse, die sie im positiven Unendlichen berühren wird.

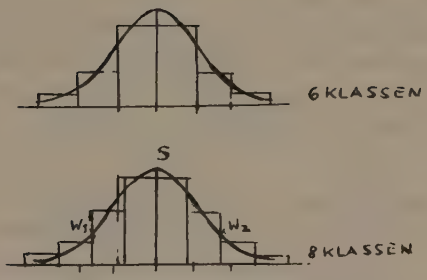


Folgende Charakteristika der Häufigkeitsverteilung liegen bei der standardisierten Normalverteilung vor:
Zwischen o und 1 z liegen 34,13% der Fälle, zwischen 1 und 2 z 13,59% und zwischen 2 und 3 z 2,15%. Zwischen 3 z und dem Berührungspunkt der Kurve mit der X-Achse liegen die restlichen o,13% der Fälle.

Eine vernünftige Klasseneinteilung muß nun so vorgenommen werden, daß durch die Einteilung die Eigenarten der Normalverteilung nicht wegfallen. Sie darf nicht so vorgenommen werden, daß durch sie eine Vielzahl von Verteilungen beschrieben werden könnte:

Wählen wir eine Einteilung in drei Klassen wie in der Abbildung oben, so kann diese Klasseneinteilung eine Normalverteilung kennzeichnen, aber auch noch andere Verteilungen; diese Einteilung ist nicht charakteristisch genug, damit sie nur für die Normalverteilung gelten könnte.

Wieviele Klassen sind nun notwendig, um eindeutig eine Normalverteilung zu beschreiben?
Geht man von den speziellen Häufigkeitsverhältnissen der Normalverteilung aus und macht jede Klasse so groß wie 1 z, so benötigt man 6 Klassen. Bei dieser Einteilung werden aber einige Besonderheiten des Kurvenverlaufs außer Acht gelassen. Eine Einteilung in acht Klassen wird diesem mehr gerecht.



Will man möglichst optimal vorgehen, so wählt man eine Einteilung in 8 Klassen. Wenn man außerdem der Forderung des Chi-Quadrats gerecht werden will, so muß in jeder Klasse mindestens eine Häufigkeit von 5 sein. Nehmen wir an, in unserer kleinsten Klasse sollen 1% der Fälle liegen, so muß dieses 1% einer Häufigkeit von 5 entsprechen. 1oo% entspricht dann einer Häufigkeit von 5oo, d.h. wir benötigen 5oo Rohwerte oder in diesem Zusammenhang einen Versuch mit 5oo Vpn. Der Aufwand für diesen Versuch ist so groß, daß er nicht mehr im Verhältnis zu dem methodischen Gewinn steht. Es ist deshalb ökonomischer, methodische Abstriche zu machen und dadurch den Versuchsaufwand zu reduzieren.

Nimmt man die Klasseneinteilung in 6 Klassen, die den Verlauf der Kurve nicht so exakt berücksichtigt, aber den Häufigkeitsverhältnissen gerecht wird, so liegen in der kleinsten Klasse 2,15% der Fälle. Wenn nach der Forderung des Chi-Quadrats diese 2,15% einer Häufigkeit von 5 entsprechen sollen, so benötigt man insgesamt 235 Rohwert oder Vpn (5/2,15 = 2,35 entspricht 1%, 1oo% ist dann 2,35 x 1oo).

Auch hier liegt der Versuchsaufwand noch sehr hoch. Da wir von der optimalen Wiedergabe der Normalverteilung durch 8 Klassen abgesehen haben, können wir nun von der Forderung des Chi-Quadrats nach mindestens 5 Häufigkeiten pro Klasse gewisse Abstriche machen. In der kleinsten Klasse sollen 2,15% der Fälle sein. Die Besetzung dieser Klasse muß nun so gewählt werden, daß diese 2,15% eine echte

Häufigkeit bedeuten. Bei insgesamt 15 Rohwerten oder Vpn sind 2% o,3 Häufigkeiten, diese Klasse fiele demnach fort. Da die Klasseneinteilung symmetrisch ist, fiele am anderen Ende die entsprechende Klasse ebenfalls fort und wir haben nur noch 4 Klassen übrig, die dann nicht mehr eine Normalverteilung repräsentieren können.

Bei der Klasseneinteilung in 6 Klassen sollte man wenigstens 1oo Rohwerte zur Verfügung haben, damit auch die Extremklassen besetzt sein können, d.h. daß sie echte Häufigkeiten enthalten. Bei insgesamt 5o Rohwerten, wären die Extremklassen mit einer Häufigkeit besetzt, wodurch dann die Gefahr sehr groß wird, daß diese Häufigkeit aufgrund von Zufallsschwankungen nicht erreicht wird, und die Klasse unbesetzt bleibt oder zufällig 2 Vpn diese Klasse wählen, wodurch die Verteilung beeinträchtigt wird.

Wenden wir nun diese Überlegungen auf die hier zur Frage stehenden Untersuchungen an, so müssen wir feststellen, daß bei allen Versuchen die Vpn-Zahlen so gering sind (maximal 25 Vpn pro Versuch), daß eine methodisch exakte Überprüfung auf Normalverteilung nicht möglich ist. Die einzelnen Versuche können auch nicht zusammengefaßt werden, um insgesamt größere Vpn-Zahlen zu erhalten, da die Versuche so angelegt sind, daß sie möglichst viele Variablen erforschen und kein Versuch dem anderen gleicht. Die Zusammenfassung bei Versuchswiederholungen ist auch nicht zulässig, da dann die einzelnen Rohwerte nicht mehr voneinander unabhängig sind, da jeweils zwei Werte von ein und derselben Vp stammen.

Für eine methodisch exakte Überprüfung auf Normalverteilung benötigt man Versuche mit ca. 1oo Vpn, im Grenzfall kann man sich auch mit 5o Vpn begnügen. Aber Versuche, deren Vpn-Zahlen unter 5o liegen, können für diese Überprüfung nicht verwendet werden.

Im Vordergrund dieser Arbeit steht nicht unbedingt die Beantwortung der Hypothesen; wo eine Beantwortung nicht möglich ist, sollen Wegweiser für spätere Untersuchungen gesetzt werden, die sich die Bearbeitung spezieller Hypothesen zur Aufgabe gemacht haben. Der Hinweis, daß Untersuchungen zur Normalverteilung mit mindestens 1oo Vpn durchzuführen sind, ist etwas dürftig, und in jedem guten Statistikbuch nachzulesen, er hat nichts mit der Bezugssystemproblematik zu tun.

Um etwas mehr Information über mögliche Probleme bei der Untersuchung dieser Fragestellung zu erhalten, wird eine Art Trendanalyse vorzunehmen sein. Statistisch und methodisch exakte Aussagen sind nicht möglich und sollen im Folgenden auch nicht gemacht werden. Es wird versucht, trotz der geringen Vpn-Zahlen einen Verteilungsvergleich mit Hilfe des Chi-Quadrates vorzunehmen. Die Ergebnisse stellen dabei, wie gesagt, keine statistisch fundierte Aussagen dar, sondern sollen nur gewisse mögliche Trends zu Tage treten lassen, die dann für die Planung weiterer Untersuchungen hilfreich sein können.

Zur Veranschaulichung der Vorgehensweise nehmen wir wiederum ein Beispiel. Wir wählen diesmal einen Versuch von HRUSCHKA mit Taschentüchern. Im Aufbau entspricht dieser Versuch dem mit den Armbanduhren, nur sollten hier quadratische Stofftücher bezüglich ihrer Größe als Damentaschentücher beurteilt werden. Für den Bereichsumfang der ersten Kategorie wurde die entsprechende Häufigkeitsverteilung ermittelt und aufgrund des Mittelwertes und der Streuung dieser Verteilung die zugehörige 'ideale' Normalverteilung berechnet. Für die Größe der einzelnen Gegebenheiten innerhalb dieser Kategorie erhält man so einmal den beobachteten Häufigkeitswert (Fo) und zum anderen den zu erwartenden Häufigkeitswert (Fe) der idealen Normalverteilung.

Tabelle 4

Bezugssystem: Taschentücher (Damen)
Beobachtete und erwartete Häufigkeiten bei der ersten Kategorie

S.L.	9	1o	11	12	13	14	15	16	17	18	19	2o	21	22	23	24	25	26	27
Fo	o	1	o	o	2	1	2	1	o	1	o	3	o	1	o	o	1	2	o
Fe	o.o	o.3	o.4	o.5	o.6	o.8	o.9	1.o	1.1	1.1	1.1	1.1	1.o	o.8	o.7	o.6	o.4	o.3	o.

Legende:

S.L.: Seitenlänge der Taschentücher
Fo : beobachtete Häufigkeit
Fe : erwartete Häufigkeit

Um für die beobachteten Häufigkeiten eine kontinuierliche Skala zu erhalten, muß eine Klasseneinteilung vorgenommen werden. Diese Einteilung wird nun nicht nach den Gesichtspunkten der Normalverteilung oder der Forderung des Chi-Quadrats vorgenommen, da dies sowieso unmöglich ist bei insgesamt 15 Vpn, sondern so, wie es am optimalsten ist, d.h. daß die höchsten Zufallswahrscheinlichkeiten herauskommen. Dieses Vorgehen ist möglich, da ja von statistischer Sicht her diese Analyse schon von sich aus unmöglich ist und es uns zum anderen nichts nützt, wenn wir durchweg signifikante Abweichungen erhalten, die selbst wieder keinen weiteren Aufschluß zulassen.

Von der Anzahl der Freiheitsgrade her ist es erforderlich, wenigstens 4 Klassen zu erhalten (die Anzahl der Freiheitsgrade ist gleich der Anzahl der Klassen minus 3).

Bei unserem Beispiel ist folgende Klasseneinteilung günstig:

Klasse	Fo	Fe	$(Fo-Fe)^2/Fe$
12 u. weniger	1	1.8	o.35
13 - 17	6	4.4	o.58
18 - 22	5	5.1	o.oo
23 - 27	3	2.2	o.29
		Chi-Quadrat =	1.22
		FG =	1

Bei dieser Klasseneinteilung erhalten wir ein Chi-Quadrat von 1.22 mit einem Freiheitsgrad, was einer Wahrscheinlichkeit von .3o bis .5o entspricht. Dieser Wahrscheinlichkeitswert besagt, daß die beobachtete Abweichung unserer Verteilung von der idealen Verteilung mit 3o- bis 5oprozentiger Wahrscheinlichkeit zufällig ist.

Die Interpretation der Chi-Quadrat-Werte geschieht in Anlehnung an GUILFORD [1o]. Die zu testende Hypothese lautet: die beobachteten Häufigkeiten stellen eine Zufallsabweichung der entsprechenden Häufigkeiten einer Normalverteilung mit gleichem Mittelwert und gleicher Standardabweichung dar. Bei Wahrscheinlichkeiten über .5o tolerieren wir die Hypothese, die beobachtete Häufigkeitsverteilung entstamme einer normalverteilten Grundgesamtheit. Bei Wahrscheinlichkeiten ab .o5 und weniger besteht ein gesicherter Unterschied, das Ergebnis kann nicht mehr im Sinne der Hypothese gesehen werden. Wahrscheinlichkeiten zwischen .5o und .o5 machen keine eindeutige Aussage. Ergebnisse, die in diesem Bereich liegen, werden als bedenklich bezeichnet. Um gewisse graduelle Unterschiede herauszuheben, wird bei dem bedenklichen Bereich noch folgende Einteilung gemacht, die sich an die Tabelle E von GUILFORD anschließt:
oberhalb von .o5, oberhalb von .1o, oberhalb von .2o und oberhalb von .3o.

In unserem Beispiel haben wir kein eindeutiges Ergebnis erzielt, die Wahrscheinlichkeit liegt oberhalb von .3o, aber noch unter .5o. Dieses Ergebnis ist bedenklich für eine Interpretation im Sinne der Hypothese, es liegt jedoch im 'oberen Bedenklichkeitsbereich'.

Die nun folgenden Analysen beruhen auf den hier im Einzelnen gezeigten Schritten. Es soll noch einmal darauf hingewiesen werden, daß sie keine methodischen Ansprüche stellen, sondern nur der Versuch einer gewissen Trendanalyse darstellen.

Zuerst werden die Ergebnisse von Versuchen mit Systemenden (2.), aufgeteilt nach Systemumfang, Bereichsumfang und Systemenden, besprochen, danach die Ergebnisse von Versuchen ohne Systemenden (3.), bei denen nur der Aspekt der Bereichsumfänge zu behandeln ist.

2. Versuche mit Systemenden

Dieser Abschnitt beschäftigt sich ausschließlich mit den Versuchen von HRUSCHKA, bei denen Systemenden vorhanden sind. Für die folgenden Analysen stehen uns insgesamt 16 Versuche zur Verfügung, die mit den Gegebenheiten Bleistifte, Taschentücher, Christbaumkugeln und Armbanduhren durchgeführt wurden. Bei diesen 16 Versuchen nehmen die Versuche mit den Bleistiften eine Sonderstellung ein, weil bei ihnen nicht beide Systemenden vorhanden sind. Die Skala der Bleistifte endete mit dem Normbleistift, es gab somit keine Bleistifte, die für das System zu groß waren. Im Kleinbereich wurden die Bleistifte so klein, daß sie von einer Reihe von Vpn nicht mehr als Bleistift akzeptiert wurden und somit der Systemanfang vorhanden ist.

Die folgende Tabelle soll einen Überblick über die einzelnen Versuche geben, wo sie bei HRUSCHKA zu finden sind, mit welchen Gegebenheiten, mit welcher vorgegebenen Kategorienzahl und mit welcher Vpn-Zahl sie durchgeführt worden sind. In den anschließenden Erörterungen wird auf diese Tabelle Bezug genommen. Die dort eingeführte Numerierung wird durchgängig beibehalten und die übrigen Merkmale der Versuche werden nicht jeweils wieder wiederholt.

Tabelle 6 gibt die errechneten Chi-Quadrat-Werte mit den zugehörigen Freiheitsgraden wieder. Dort sind die Chi-Quadrat-Werte für alle drei Größen, Bereichsumfang, Systemumfang und Systemenden aufgezeichnet. Bei den Bleistiftversuchen ist kein Systemende vorhanden, die zugehörige Spalte bleibt deshalb leer.

Übersicht über die Versuchsunterschiede bei HRUSCHKAs Versuchen mit Systemenden

	Gegebenheit	lfd.Nr.	Seite bei HRUSCHKA	Kategorienanzahl	Vpn-Anzahl	
a. Versuche mit einem Systemende	Bleistifte	1.	31	3	16	
		2.	46	4	15	
		3.	2o	5	25	
		4.	6o	5	15	
b. Versuche mit beiden Systemenden	Taschentücher (Damen)	1.	97	5	15	
		2.	132	3	15	
		3.	54	3	25	
		4.	91	3	15	
	Christbaumkugeln	1.	124	2	15	
		2.	127	5	18	
		3.	142	3	15	
		4.	119	3	25	
	Armbanduhren	1.	1o8	5	15	Herrenuhren
		2.	112	3	25	Damenuhren
		3.	1o5	3	25	Herrenuhren
		4.	137	3	15	Damen- und Herrenuhren

Tabelle 6

Chi-Quadrat-Werte und Freiheitsgrade bei den Versuchen von HRUSCHKA

Versuch	relative Bereichsumfänge 1.Kat.	2.Kat.	3.Kat.	4.Kat.	5.Kat.	absoluter Systemumfang	Systemenden Anfang	Ende
Taschentücher								
1.	1.22 [1]	2.15 [1]	2.66 [2]	o.67 [2]	o.81 [2]	1.21 [1]	1o.o8 [11]	8.o9 [15]
2.	2.29 [1]	o.82 [1]	o.47 [1]			o.99 [1]	8.44 [11]	2.55 [4]
3.	o.94 [1]	o.75 [1]	26.54 [25]			o.44 [1]	12.93 [13]	2.75 [4]
4.	34.54 [24]	1.79 [2]	1.13 [1]			2.2o [1]	6.18 [1]	1.o6 [2]
Christbaumkugeln								
1.	3.17 [3]	5.38 [4]				o.63 [2]	2.81 [2]	5.9o [5]
2.	1.73 [1]	2.36 [4]	3.83 [1]	1.99 [4]	o.7o [1]	14.98 [16]	34.78 [6]	16.58 [1o]
3.	14.44 [19]	13.o7 [17]	34.2o [21]			16.31 [16]	4.37 [1]	17.12 [16]
4.	19.83 [25]	o.72 [1]	1.oo [1]			2o.12 [16]	33.9o [5]	1.83 [2]
Armbanduhren								
1.	26.82 [24]	o.34 [1]	o.91 [1]	15.89 [17]	o.85 [1]	o.57 [1]	7.55 [7]	17.o2 [13]
2.	o.69 [2]	o.69 [3]	1.19 [1]			1.22 [2]	-----	3.16 [4]
3.	o.79 [1]	o.28 [1]	1.17 [3]			1.41 [1]	o.77 [1]	o.62 [1]
4.	22.23 [23]	2.39 [2]	5.o8 [5]			o.23 [1]	16.66 [12]	o.44 [1]
Bleistifte								
1.	4.1o [1]	25.13 [22]	5.24 [4]			6.91 [5]	6.67 [5]	
2.	3.8o [3]	1.3o [1]	2.63 [2]	16.25 [19]		11.76 [3]	12.o6 [3]	
3.	1.76 [2]	26.85 [22]	o.24 [1]	o.7o [1]	o.94 [2]	14.o8 [3]	1o.54 [3]	
4.	14.59 [16]	2o.o4 [23]	1.25 [1]	16.94 [19]	1.o7 [1]	5.7o [5]	5.55 [5]	

2.11 [2] bedeutet: Chi-Quadrat = 2.11 FG = 2
wurde die Zahl zusätzlich unterstrichen, so liegt an dieser Stelle eine Häufung bei o vor, das Systemende wurde nicht wahrgenommen.

Die Wahrscheinlichkeitswerte zu den Chi-Quadraten der Tabelle 6 entsprechend Tabelle E von GUILFORD

Versuch	relative Bereichsumfänge 1.Kat.	2.Kat.	3.Kat.	4.Kat.	5.Kat.	absoluter Systemumfang	Systemenden Anfang	Ende
Taschentücher								
1.	.2o	.1o	.2o	.7o	.5o	.2o	.5o	.9o
2.	.1o	.3o	.3o			.3o	.5o	.5o
3.	.3o	.3o	.3o			.5o	.3o	.5o
4.	.o5	.3o	.2o			.1o	*.o1*	.5o
Christbaumkugeln								
1.	.3o	.2o				.7o	.2o	.3o
2.	.1o	.5o	.o5	.7o	.3o	.5o	*unter .oo1*	.o5
3.	.7o	.7o	*.o2*			.3o	*.o2*	.o5
4.	.7o	.3o	.3o			.2o	*unter .oo1*	.3o
Armbanduhren								
1.	.3o	.5o	.3o	.5o	.3o	.3o	.3o	.1o
2.	.7o	.8o	.2o			.5o	---	.5o
3.	.3o	.5o	.7o			.2o	.3o	.3o
4.	.5o	.3o	.3o			.5o	.1o	.5o
Bleistifte								
1.	*.o2*	.2o	.2o			.2o	.2o	
2.	.2o	.2o	.2o	.5o		*.oo1*	*.oo1*	
3.	.3o	.2o	.5o	.3o	.5o	*.oo1*	*.o1*	
4.	.5o	.5o	.2o	.5o	.3o	.3o	.3o	

Wurde eine Zahl zusätzlich unterstrichen, so liegt an dieser Stelle eine Häufung bei o vor, das Systemende wurde nicht wahrgenommen.

Tabelle 6 gibt alle berechneten Chi-Quadrat-Werte wieder und jeweils hinter dem Wert, etwas hochgesetzt, den zugehörigen Freiheitsgrad. Schlägt man die zugehörigen Wahrscheinlichkeitswerte nach, so erhält man die konkrete Information in Bezug auf die zu testende Hypothese der Normalverteilung. In Tabelle 7 sind alle diese Wahrscheinlichkeitswerte zusammengefaßt aufgeführt. Sie wurden der Tabelle E von GUILFORD entnommen.

Wahrscheinlichkeiten, die gleich oder größer .5o sind, und Wahrscheinlichkeiten, die kleiner .o5 sind, machen eine eindeutige Aussage bezüglich der Hypothese. Wahrscheinlichkeiten, die zwischen diesen beiden Werten liegen, sind nicht eindeutig, wir haben sie mit bedenklich bezeichnet und diesen Bereich noch unterteilt (siehe Seite 31).

In Tabelle 7 streuen die einzelnen Wahrscheinlichkeiten recht stark. Für einen generellen Überblick über die Wahrscheinlichkeitsverhältnisse ist es zunächst angebracht, die Häufigkeit der einzelnen Wahrscheinlichkeitswerte auszuzählen. In Tabelle 8 ist das Ergebnis dieser Auszählung wiedergegeben.

Tabelle 8

Häufigkeit der Wahrscheinlichkeiten von Tabelle 7

	Größe der Wahrscheinlichkeitswerte						
	über .5o	.3o-.5o	.2o-.3o	.1o-.2o	.o5-.1o	unter .o5	Ges
f	33	3o	18	6	4	11	1o2
%	33	29	18	6	4	1o	1oo

Legende:
f : Häufigkeit
% : in Prozentzahlen ausgedrückt
Ges: Summe der Häufigkeiten oder Prozentzahlen

Insgesamt gesehen liegen 33% der Wahrscheinlichkeiten über .5o und nur 1o% davon unter .o5. Die Mehrzahl der bedenklichen Ergebnisse liegt im Bereich von .3o - .5o, also im oberen 'Bedenklichkeitsbereich'.

Im Folgenden werden die Ergebnisse von Tabelle 6 und 7 für die einzelnen Größen (relativer Bereichsumfang, absoluter Systemumfang und Systemenden) aufgeschlüsselt und behandelt. Begonnen wird mit den Systemenden (2.1), dann folgt der absolute Systemumfang (2.2) und die relativen Bereichsumfänge (2.3).

a. Prüfung auf Normalverteilung bei Systemanfang und Systemende

Hypothese 2 besagt, daß Systemanfang und -ende normalverteilt seien. D.h.: Die Anzahl der Gegebenheiten, die vor oder hinter einem System liegen, oder die nichtkategorialen Beurteilungen entsprechen in ihrer Verteilung einer Normalverteilung.

Für die Prüfung dieser Hypothese wurde die Häufigkeitsverteilung dieser Größen und die zugehörige Normalverteilung mit gleichem Mittelwert und gleicher Standardabweichung berechnet. Beide Verteilungen wurden mittels des Chi-Quadrats miteinander verglichen und die Wahrscheinlichkeit einer zufälligen Abweichung der beobachteten Verteilung von der Normalverteilung berechnet. Die Ergebnisse bezüglich des Systemanfanges und -endes sind in Tabelle 6 (Seite 35) und Tabelle 7 (Seite 36) aufgezeigt. Greift man aus diesen beiden Tabellen die Ergebnisse bezüglich des Systemanfanges und -endes heraus, so erhält man die folgende Tabelle (Tabelle 9, nächste Seite).

Bei dieser Tabelle sind die Häufigkeiten der einzelnen Wahrscheinlichkeitswerte von Systemanfang und Systemende herausgegriffen, u.zw. getrennt nach Versuchen mit beiden Systemenden und den Bleistiftversuchen mit nur einem Systemende (siehe Seite 33).

Betrachten wir zuerst die Versuche mit beiden Systemenden, so sind die Verteilungen beim Systemende recht hypothesengerecht ausgefallen. 5o% der Ergebnisse bestätigen die Hypothese, sie liegen über .5o, kein Ergebnis widerlegt die Hypothese signifikant. Von den restlichen 5o% der Ergebnisse liegen die Hälfte (25%) im Bereich .3o - .5o, d.h. im 'oberen Bedenklichkeitsbereich'.

Die Verteilungen beim Systemanfang haben nur wenig Ähnlichkeit mit einer Normalverteilung. 42% der Ergebnisse zeigen einen signifikanten Unterschied an und nur 17% bestätigen die Hypothese. Wie bei den System-

enden liegen wiederum 25% der Ergebnisse im 'oberen Bedenklichkeitsbereich', also im Bereich von .3o - .5o.

Tabelle 9

Häufigkeiten der einzelnen Wahrscheinlichkeitswerte bei Systemanfang und Systemende, entnommen Tabelle 7

Wahrscheinlichkeitswerte	Systemanfang a. mit beiden Systemenden	Systemende a. mit beiden Systemenden
über .5o	2 = 17%	6 = 5o%
.3o - .5o	3 = 25%	3 = 25%
.2o - .3o	1 = 8%	o = o%
.1o - .2o	1 = 8%	1 = 8%
.o5 - .1o	o = o%	2 = 17%
unter .o5	5 = 42%	o = o%
	b. mit einem Systemende	
über .5o	o = o%	
.3o - .5o	1 = 25%	
.2o - .3o	1 = 25%	
.1o - .2o	o = o%	
.o5 - .1o	o = o%	
unter .o5	2 = 5o%	

Bei den Bleistiftversuchen mit nur einem Systemende fallen die Ergebnisse noch ungünstiger aus. 5o% widersprechen der Hypothese eindeutig, 5o% liegen im Bedenklichkeitsbereich. Kein Ergebnis liegt über .5o.

b. Prüfung auf Normalverteilung beim absoluten Systemumfang

Hypothese 1 besagt, daß der absolute Systemumfang normalverteilt sei. Die Größe der persönlichen Systeme der einzelnen Vpn sollen in ihrer Verteilung einer Normalverteilung entsprechen.

Es wurden wiederum die beobachteten Verteilungen und die zugehörigen Normalverteilungen berechnet und dann miteinander verglichen. Die Chi-Quadrate dieser Verteilungen und die zugehörigen Wahrscheinlichkeitswerte sind in Tabelle 6 (Seite 35) und Tabelle 7 (Seite 36) nachzusehen.

Die sich auf den Systemumfang beziehenden Ergebnisse der beiden Tabellen werden hier noch einmal herausgegriffen und die Häufigkeiten der einzelnen Wahrscheinlichkeiten für den Systemumfang aufgezeigt, wobei zwischen den Versuchen mit beiden Systemenden und den Bleistiftversuchen mit nur einem Systemende unterschieden wird.

Tabelle 1o

Häufigkeit der einzelnen Wahrscheinlichkeitswerte beim absoluten Systemumfang, entnommen Tabelle 7

Wahrscheinlichkeitswerte	a. mit beiden Systemenden	b. mit einem Systemende
über .5o	5 = 42%	o = o%
.3o - .5o	3 = 25%	1 = 25%
.2o - .3o	3 = 25%	1 = 25%
.1o - .2o	1 = 8%	o = o%
.o5 - .1o	o = o%	o = o%
unter .o5	o = o%	2 = 5o%

Die Ergebnisse bei den Bleistiftversuchen mit nur dem Systemanfang sind völlig identisch mit den Ergebnissen des Systemanfanges bei diesen Versuchen.

Die Hälfte der Ergebnisse zeigt einen gesicherten Unterschied zur Normalverteilung an, die andere Hälfte ergibt nur bedenkliche Wahrscheinlichkeiten.

Bei den Versuchen mit beiden Systemenden liegt die Verteilung des absoluten Systemumfanges wesentlich besser. Es gibt keine Wahrscheinlichkeit unter .1o und 42% liegen über .5o und entsprechen der Hypothese.

c. Prüfung auf Normalverteilung bei den relativen Bereichsumfängen

Die relativen Bereichsumfänge, d.h. die auf den absoluten Systemumfang bezogenen Werte der einzelnen Bereiche, sollen ebenfalls normalverteilt sein (Hypothese 1).

Die Ergebnisse zu den entsprechenden Berechnungen sind in Tabelle 6 (Seite 35) und Tabelle 7 (Seite 36) zu finden. Hier sollen die Ergebnisse bezüglich der relativen Bereichsumfänge nochmals aufgezeigt werden. Es wird dabei unterschieden zwischen Versuchen mit nur einem Systemende und Versuchen mit beiden Systemenden. Die Ergebnisse aller Kategorien werden zusammengefaßt und die Häufigkeiten der einzelnen Wahrscheinlichkeitswerte in der nachfolgenden Tabelle aufgeführt.

Tabelle 11

Häufigkeiten der einzelnen Wahrscheinlichkeitswerte bei den relativen Systemumfängen, entnommen Tabelle 7

Wahrscheinlichkeitswerte	a. mit beiden Systemenden	b. mit einem Systemende
über .5o	14 = 34%	6 = 35%
.3o - .5o	16 = 39%	3 = 18%
.2o - .3o	5 = 12%	7 = 41%
.1o - .2o	3 = 8%	o = o %
.o5 - .1o	2 = 5%	o = o %
unter .o5	1 = 2%	1 = 6%

Hier liegen prozentual gesehen bei den Versuchen mit beiden Systemenden und denen mit einem Systemende gleich viele Ergebnisse über .5o und entsprechen der Hypothese. Auch bei den niedrigeren Wahrscheinlichkeiten sind die Unterschiede nicht sehr groß. Bei den Versuchen mit beiden Systemenden liegen die meisten Ergebnisse im Bereich .3o - .5o, bei denen mit einem Systemende im Bereich .2o - .3o.

2. Versuche ohne Systemenden

Das Datenmaterial der Versuche ohne Systemenden kann nur für die Bearbeitung der relativen Bereichsumfänge herangezogen werden, da die anderen Größen bei diesen Versuchen nicht vorhanden sind (siehe Seite 17, 18). Die Ergebnisse bei diesen Versuchen bezüglich der Bereichsumfänge einzeln aufzuzeigen, bringt wenig Aufschluß. Sie sind sehr umfangreich und unterscheiden sich im Wesentlichen nicht von den Ergebnissen bezüglich der Bereichsumfänge bei Versuchen mit Systemenden. Die Verteilung der Ergebnisse, die im Sinne der Hypothese ausfallen oder diese widerlegen, entspricht den schon behandelten.

II. Diskussion der Ergebnisse

Betrachtet man die Gesamtheit der Ergebnisse bei der Prüfung auf Normalverteilung, so können drei Punkte hervorgehoben werden:

1. Bei den Systemenden unterscheiden sich die Ergebnisse von Systemanfang und Systemende.

2. Beim absoluten Systemumfang unterscheiden sich die Ergebnisse von den Versuchen mit beiden Systemenden von denen mit nur einem Systemende.

3. Die Ergebnisse der relativen Bereichsumfänge zeigen keine großen Unterschiede, ob es sich nun um Versuche mit beiden Systemenden, mit nur einem oder gar keinem Systemende handelt.

Diesen unterschiedlichen Ergebnissen bei ein und derselben Größe müssen Variablen zugrundeliegen, die in den einzelnen Versuchen zu suchen sind. Wären die Ergebnisse bei Systemanfang und Systemende entweder gleich gut oder schlecht ausgefallen, so hätte man darüber keine weiteren Aussagen machen können (da die hier durchgeführte rechnerische Prozedur keinen Anspruch auf statistisch gesicherte Aussagen erheben kann). Unterschiede bei den Ergebnissen können aber trotz der methodischen Unzulänglichkeiten Aufschlüsse geben über gewisse Beeinflussungsverhältnisse, die bei späteren Untersuchungen zu beachten sind.

Die Ergebnisse der relativen Bereichsumfänge sind bei allen Versuchsarten recht gleichbleibend. Sie liegen insgesamt gesehen für die Hypothese ganz gut, was aber nicht zu der Aussage verleiten darf, daß die Normalverteilung der relativen Bereichsumfänge dadurch gesichert sei. Man kann lediglich sagen, daß die Versuchsart, Versuche mit oder ohne Systemenden, keinen Einfluß zu haben scheint. Die anfangs geäußerte

Vermutung, daß die Bereichseinteilung durch die Einschränkung auf den für alle Vpn akzeptablen Mittelbereich ihrer Systeme beeinflußt werden könnte (siehe Seite 18), hat sich nicht bestätigt.

Aufschlußreicher wird es sein, die Unterschiede bei den Systemenden und dem absoluten Systemumfang näher zu betrachten, da sie Anhaltspunkte für Versuchsplanungen liefern können.

1. Unterschied zwischen Systemanfang und -ende

Die Ergebnisse beim Systemende sind von der Hypothese aus gesehen ganz gut. Die Hälfte der Ergebnisse liegt über .5o. Beim Systemanfang sind die Verhältnisse fast umgekehrt, die Hälfte der Ergebnisse liegt unter .o5. Um diesen Gegensatz zu klären, erscheint es angebracht, die Ausgangsdaten noch einmal zu untersuchen. Bei den Ergebnissen des Systemanfanges ist in Tabelle 6 (Seite 35) und Tabelle 7 (Seite 36) mehrmals durch Unterstreichen gekennzeichnet, daß hier bei den Verteilungen eine Häufung bei Null vorliegt. Das bedeutet für die Versuche, daß die Vpn im Kleinbereich oft alle Gegebenheiten akzeptiert haben und nicht eine so klein war, daß sie nicht mehr in das System der Vp paßte.

Den Ergebnissen des Systemanfangs liegen vier Versuchsarten zugrunde: Versuche mit Taschentüchern, mit Christbaumkugeln, mit Armbanduhren und mit Bleistiften. Zu jeder Versuchsart liegen die Ergebnisse von vier Einzelversuchen vor. Bei den vier Einzelversuchen zu einer Gegebenheit wurde die Skala der Variationsbreite nicht verändert. Dies ist von Bedeutung, da bei den Armbanduhren die dargebotenen Uhren einmal von Herren als Herrenuhren, zum andern von Damen als Damenuhren beurteilt werden

sollten. Der vierte Einzelversuch bei den Uhren stellt noch eine Vereinigung dieser beiden Versuchsvariationen dar, indem hier die Urteile von 8 Herren und 7 Damen zusammengefaßt sind. Ein geschlechtsspezifischer Unterschied wurde auch bei den Taschentüchern gemacht, nur liegen uns hier nur die Ergebnisse von Damenbeurteilungen bezüglich Damentaschentüchern vor. Die Christbaumkugeln und Bleistifte wurden von Damen und Herren beurteilt.

Bei den 16 Versuchen wurde nun ausgezählt, wieviele Vpn alle Gegebenheiten im Kleinbereich akzeptierten, d.h. wie oft kein Systemanfang vorhanden war. Diese Auszählung wurde auch für die Systemenden vorgenommen, wobei die Bleistiftversuche entfallen, da hier keine Systemenden vorhanden sind. Die folgende Tabelle zeigt das Ergebnis der Auszählung für jeden einzelnen Versuch.

Tabelle 12

Häufigkeit des Nicht-Verwendens der einzelnen Systemenden

Versuch:	T.T.				Ch.K.				A.U.				B.			
Nr. :	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
S.A.	1	o	o	1	4	11	7	12	o	9	o	o;4	2	9	4	2
S.E.	o	o	o	o	o	o	o	o	o	o	o	o;o	-	-	-	-
Vpn	15	15	25	15	15	18	15	25	15	25	25	8;7	16	15	25	15

Legende:

T.T.	: Taschentücher	S.A.	: Systemanfang
Ch.K.	:Christbaumkugeln	S.E.	: Systemende
A.U.	: Armbanduhren	Vpn	: Anzahl der Vpn
B.	: Bleistifte		

Beim vierten Uhrenversuch ist zwischen Damen- und Herrenbeurteilung unterschieden. Links vom Semikolon stehen die Werte der Herren- und rechts davon die der Damenbeurteilungen. Versuch 1 und 3 wurde von Herren und Versuch 2 von Damen alleine beurteilt.

Sehr auffällig ist die Tatsache, daß bei den Systemenden nie der persönliche Bereich einer Vp bis zum Ende der vorgegebenen Skala reichte, daß die Skala also im Großbereich so gewählt worden war, daß die extremen Gegebenheiten für alle Vpn nicht mehr akzeptabel waren.

Im Kleinbereich reichte die Skalenlänge oft nicht aus und es wurden viele Gegebenheiten am Ende der Skala noch akzeptiert. Es ist klar, daß bei diesen Verhältnissen kaum eine Verteilung entstehen kann, die Normalverteilungscharakter besitzt. Bei einer Normalverteilung könnte Null nur die äußerste Extremklasse darstellen, da es negative Zahlen hier nicht geben kann. Ist nun diese Extremklasse häufig besetzt, kann keine Normalverteilung mehr herauskommen.

Bei den Christbaumkugel- und Bleistiftversuchen sind die Extremklassen beim Systemanfang so häufig besetzt, daß es erstaunlich ist, daß teilweise noch Ergebnisse um .2o und .3o herauskommen. Diese Ergebnisse wurden jedoch bei den Versuchen erzielt (Versuch 1 der Christbaumkugeln und Versuch 1 und 4 der Bleistifte), bei denen im Verhältnis zu den anderen Versuchen der Systemanfang nur selten nicht benutzt wurde (4mal, 2mal, 2mal). Bei diesen Gegebenheiten wurde die Skalenbreite im Kleinbereich einfach zu klein gewählt, um echte Systembegrenzungen zu erhalten.

Bei den Taschentüchern, bei denen die Ergebnisse auch beim Systemanfang noch recht akzeptabel sind, reicht die Skalenlänge nur vereinzelt nicht aus. Bei den Armbanduhren muß man zwischen den Beurteilungen auf Herrenuhren und Damenuhren hin unterscheiden. Für die Herrenuhren reichte die Skalenlänge auch im Kleinbereich aus, die dargebotenen extrem kleinen Uhren sind durchgängig für Herrenuhren zu klein. Die Ergebnisse des

Systemanfanges bei den Herrenbeurteilungen sind auch gegenüber den Damenbeurteilungen besser. Das Ergebnis des vierten Versuchs kann man als ein Mittel verstehen, daß aus einem ganz guten Ergebnis der Herren und einem schlechten der Damen zusammengesetzt ist. Für die Beurteilungen auf Damenuhren hin reichte die Skalenlänge nicht aus, viele der ganz kleinen dargebotenen Uhren wurden noch als Damenuhren akzeptiert.

Der Unterschied der Ergebnisse bei Systemanfang und Systemende läßt sich auf unterschiedliche Bedingungen zurückführen:
Im Kleinbereich waren die dargebotenen Gegebenheiten oft nicht klein genug, um nicht mehr akzeptiert zu werden, und es gab somit keinen Systemanfang für einzelne Bezugssysteme.

2. Absoluter Systemumfang bei Versuchen mit einem und mit beiden Systemenden

Die Ergebnisse des absoluten Systemumfanges bei Versuchen mit beiden Systemenden liegen recht gut, fast die Hälfte über .5o und keines unter .o5. Davon heben sich die Ergebnisse bei den Bleistiftversuchen mit nur einem Systemende ab. Die Ergebnisse des Systemumfanges entsprechen hier fast völlig denen des Systemanfanges. Diese Gleichheit der Ergebnisse leitet sich aus dem nur am Anfang der Skala vorhandenen Systemende ab. Fehlt das zweite begrenzende Systemende, so sind die Werte von Umfang und Ende direkt voneinander abhängig. Die Gesamtbreite der Skala ist die Summe von Systemanfang und -umfang. Ist der Wert des Systemanfanges bekannt, ergibt sich daraus der Wert des Systemumfanges. Entspricht die Verteilung des Systemanfanges nicht einer Normalverteilung, kann die des Umfanges auch nicht einer Normalverteilung entsprechen. Die

zusammengehörigen Verteilungen müssen gleiche Werte bei der Prüfung auf Normalverteilung erzielen.

Die unterschiedlichen Ergebnisse bei dem Systemumfang von Versuchen mit beiden Systemenden und denen mit nur einem Systemende beruhen auf der selben Tatsache wie die unterschiedlichen Ergebnisse bei Systemanfang und -ende. Durch das häufige Nicht-Verwenden des Systemanfanges bei den Bleistiftversuchen können die Ergebnisse nicht einer Normalverteilung entsprechen. Das häufige Nicht-Verwenden des Systemanfanges bei den Versuchen mit beiden Systemenden wirkt sich auf den Systemumfang deshalb nicht aus, weil bei diesen Versuchen keine direkte Abhängigkeit zwischen dem Wert des Systemanfanges und -umfanges besteht, da hier noch die Größe des Systemendes hinzukommt.

III. Zusammenfassung der Ergebnisse und Folgerungen für spätere Untersuchungen

In diesem Abschnitt standen die beiden ersten Hypothesen zu Diskussion:

1. Die relativen Bereichsumfänge sowie der absolute Systemumfang sind normalverteilt.
2. Die Systemenden sind normalverteilt.

Die Verteilung der relativen Bereichsumfänge, des absoluten Systemumfanges und der Systemenden muß auf Normalverteilung überprüft werden. Diese Überprüfung sollte anhand bereits gemachter Versuche durchgeführt werden. Notwendig dafür sind Versuche, bei denen die Bezugssysteme der einzelnen Vpn sich nicht über die gesamte dargebotene Skala erstrecken, sondern am Anfang und Ende der Skala Gegebenheiten liegen, die nicht mehr in die Systeme passen. Diese Bedingung muß für die Überprüfung der relativen Bereichsumfänge nicht erfüllt sein.

Aufgrund der Forderung nach echten Systembegrenzungen blieben als Versuchsmaterial nur die Versuche von HRUSCHKA übrig. Für die Überprüfung einer Verteilung auf Normalverteilung werden von statistischer Seite her gewisse Forderungen an die Verteilung gestellt. Die Häufigkeitsverteilung darf keine Lücken aufweisen,d.h. es darf innerhalb der kontinuierlich ansteigenden Skala keine Häufigkeit von Null vorkommen. Um diese Möglichkeit auszuschalten, teilt man die Skala in Klassen ein. An die Klasseneinteilung stellt die Normalverteilung und das Chi-Quadrat gewisse Forderungen; die Einteilung muß so vorgenommen werden, daß sie die Eigenschaften der Normalverteilung wiedergibt, und muß eine sinnvolle geringste Besetztheit aufweisen.

Keine der zur Verfügung stehenden Verteilungen wird den statistischen Anforderungen gerecht. Die Versuche, denen das Datenmaterial entstammt, wurden im Rahmen einer empirischen Erkundung durchgeführt und nicht in

Hinblick auf statistische Verarbeitung.

Um diese Fragestellung nicht völlig unbearbeitet zur Seite legen zu müssen, wurde eine Art Trendanalyse vorgenommen. Das vorhandene Datenmaterial wurde auf Normalverteilung geprüft, als wenn es keine Einschränkungen von methodischer Seite her gäbe. Hiermit wurde nicht eine methodisch unexakte Aussage über Normalverteilung oder nicht angestrebt. Die Analyse sollte mögliche Hinweise über Beeinflussungsverhältnisse aufdecken, die für spätere Versuchsplanungen von Wichtigkeit sein können. Die erzielten Chi-Quadrat- und Wahrscheinlichkeitswerte, die bei der Verrechnung herausgekommen sind, sind von statistischer Seite her gesehen völlig aussagefrei, sie können und sollen nicht die Hypothese bestätigen oder widerlegen.

Das Ergebnis dieser Trendanalyse hebt zwei Punkte hervor:

1. Die Verteilungen des relativen Bereichsumfanges werden von der Versuchsart nicht beeinflußt, sie sind gleichbleibend, ob ihre Daten nun Versuchen mit beiden Systemenden oder Versuchen mit einem oder keinem Systemende entstammen.

2. Die Verteilungen der Systemenden und des Systemumfanges reagieren empfindlich auf nichtvorhandene Systemenden. Bei Versuchen mit nur einem Systemende entsprechen die Verteilungen des Systemumfanges denen des Systemendes, da beide direkt von einander abhängig sind. Bei Versuchen mit beiden Systemenden bleibt der Systemumfang unbeeinflußt von teilweise nicht vorhandenen Systemenden, da hier keine direkte Abhängigkeit mehr zwischen den Größen besteht. Die Verteilung der Systemenden selbst wird jedoch davon beeinflußt, ob das jeweilige Systemende bei einer Vielzahl von Vpn nicht verwendet wurde.

Für spätere Versuchsplanungen sind aufgrund dieser Untersuchung zwei Punkte zu beachten:

1. Die Stichprobengröße

Bei Versuchen, die speziell für die Überprüfung auf Normalverteilung angelegt werden sollen, müssen die Vpn-Zahlen gegenüber den bisherigen Versuchen erheblich erhöht werden. Die statistischen Forderungen bezüglich der Klasseneinteilung legen Vpn-Zahlen von 1oo pro Versuch (siehe Seite 25 ff) nahe. In Einzelfällen können auch 5o Vpn ausreichend sein, von der Planung her sollte man jedoch mit 1oo Vpn rechnen.

Anforderungen an die Stichprobenauswahl sind nicht zu stellen, da hier allgemein-menschliche Merkmale untersucht werden sollen und die Gegebenheiten für jeden Menschen beurteilbar sind.

2. Der Versuchsaufbau

Der Versuchsaufbau selbst wird einige Schwierigkeiten mit sich bringen. Die Skala der Gegebenheiten muß so groß gewählt werden, daß echte Systembegrenzungen bei allen Vpn vorhanden sind. Bei einem Versuch mit 5o Vpn kann bereits das zweimalige Nicht-Verwenden des Systemendes die Verteilung beeinflussen, da die Extremklasse somit eine Häufigkeit von 2 besitzt, jedoch nur eine Häufigkeit von 1 besitzen darf. Je größer die Vpn-Zahl bei dem Versuch, desto unempfindlicher werden die Verteilungen gegenüber dem Nicht-Verwenden von Systemenden. Die Variationsbreite der Gegebenheiten sollte jedoch so gewählt sein, daß nur in seltenen Ausnahmefällen das System einer Vpn noch über die Grenzen hinausgeht.

Mit Hilfe von Vorversuchen kann man den Variationsbereich abstecken. Sollte es sich herausstellen, daß die gewählte Gegebenheit nicht weiter zu variieren ist, so sollte man eine andere Gegebenheit wählen.

Hat man einen Versuch durchgeführt, bei dem durchgängig echte Systembegrenzungen vorhanden sind, so kann man mit den Daten dieses Versuches drei Fragen beantworten:

(1) Sind die relativen Bereichsumfänge, (2) sind die Systemumfänge und (3) sind die Systemenden normalverteilt?

Die Frage nach der Normalverteilung der Bereichsumfänge kann auch anhand anderer Versuche ohne Systemenden beantwortet werden, es ist aber ökonomischer, Versuche zu planen, bei denen alle drei Fragen beantwortet werden können. Sollte sich im Laufe der Versuchsdurchführung zeigen, daß keine echten Systembegrenzungen vorhanden sind, so kann man dann das Material dieses Versuches für die Beantwortung der Frage nach Normalverteilung der Bereichsumfänge heranziehen und die beiden anderen Fragen außer Acht lassen.

B. Prüfung auf Unabhängigkeit

I. Die statistische Bearbeitung

Unabhängigkeit wird in den folgenden Hypothesen angenommen:

3. Jeder relative Bereichsumfang ist unabhängig vom absoluten Systemumfang.
4. Systemanfang und -ende sind unabhängig von einander.
5. Systemanfang und -ende sind unabhängig von den relaiven Bereichsgrößen.

Die Hypothesen besagen, daß jeweils zwei Größen nicht voneinander abhängen:

(1) der relative Bereichsumfang nicht vom Systemumfang,
(2) der Systemanfang nicht vom Systemende und
(3) Systemanfang und -ende nicht von den relativen Bereichsgrößen.

Die statistische Bearbeitung beinhaltet somit die Überprüfung dieser Paare auf Unabhängigkeit, die anhand des Datenmaterials von HRUSCHKA vorgenommen werden kann. Bei diesen Versuchen wurde mit Systemenden gearbeitet, die für die Überprüfung jedes Paares erforderlich sind. Die Versuchsangaben und deren Numerierung, die in Tabelle 5, Seite 34 aufgeführt sind, treffen für die hier anstehende Untersuchung ebenfalls zu.

1. Einleitende Überlegungen

Für die Prüfung auf Abhängigkeit steht die Korrelationsrechnung zur Verfügung. Man könnte wohl versucht sein, jedesmal, wenn so keine Abhängigkeit herauskommt, Unabhängigkeit anzunehmen. Damit beginge man aber einen

statistisch bedeutsamen Fehler, der in der statistischen Handhabung der Hypothesen und deren Testung begründet liegt. Es ist deshalb notwendig, auf die Elemente der statistischen Entscheidung näher einzugehen:

Eine empirische Untersuchung beginnt meistens damit, daß wir eine bestimmte Vermutung über einen Zusammenhang haben. Diese Vermutung wollen wir dann mit Hilfe eines Experimentes überprüfen. Unsere Vermutung formulieren wir dann in Form einer Hypothese, die durch den Versuch bestätigt oder widerlegt werden soll. Diese Hypothese wird mit Arbeitshypothese oder H_1 bezeichnet und geht aus der Versuchsfrage hervor. Dieser Hypothese wird nun ihre Negation, die Nullhypothese oder H_o entgegengesetzt, die dazu dient, die Versuchsergebnisse gegen den Zufall abzusichern.

Bei der Korrelationsrechnung liegt die Vermutung zugrunde, daß ein Zusammenhang zwischen zwei Größen bestehe. Diese Vermutung wird in der Arbeitshypothese ausgedrückt und das Gegenteil in der Nullhypothese, die dann lauten:

H_1 : Es besteht ein Zusammenhang zwischen A und B.
H_o : Es besteht kein Zusammenhang zwischen A und B.

Das Ergebnis einer statistischen Analyse ist ein Wahrscheinlichkeitswert, der die Wahrscheinlichkeit der Nullhypothese angibt. Hat die Nullhypothese eine sehr geringe Wahrscheinlichkeit, so werden wir diese verwerfen und die Arbeitshypothese unter Vorbehalt annehmen. Die dabei jedoch vorhandene Wahrscheinlichkeit der Nullhypothese ist die Irrtumswahrscheinlichkeit, die wir in Kauf zu nehmen gewillt sind.
Die statistische Entscheidung beinhaltet nur die Wahrscheinlichkeit der Nullhypothese, die Irrtumswahrscheinlichkeit, die wir akzeptieren, wenn wir die Arbeitshypothese annehmen. Es wird keine Aussage über die Wahrscheinlichkeit der Arbeitshypothese selbst gemacht.

Das Ergebnis bei einer Korrelationsrechnung ist ein Koeffizient r, der Ausdruck für den Grad des Zusammenhanges sein soll. Der Wert des Koeffi-

zienten bewegt sich innerhalb der Grenzen + 1 und - 1, wobei Null keinen Zusammenhang bedeutet. Ob ein Zusammenhang statistisch gesichert ist oder nicht, hängt einmal von der Größe des Koeffizienten und zum anderen von der Anzahl der Freiheitsgrade ab. Es gibt nun Tabellen, die diese beiden Angaben berücksichtigen und anzeigen, ob ein bestimmter Koeffizient auf dem 1%- oder 5%-Niveau signifikant ist, d.h. daß die Nullhypothese hier eine Wahrscheinlichkeit von 1% oder 5% hat, die wir als Irrtumswahrscheinlichkeit in Kauf nehmen wollen. Die r-Werte dieser Tabellen stellen somit Grenzwerte dar; liegt der gefundene Wert darüber, wird die Arbeitshypothese beibehalten, liegt er darunter, wird die Nullhypothese beibehalten.

Bei unserer Aufgabenstellung geht nun die Vermutung nicht dahin, daß Zusammenhang besteht, sondern umgekehrt, daß kein Zusammenhang besteht. Die Hypothesen müssen darum umgekehrt formuliert werden:

H_1 : Es besteht kein Zusammenhang zwischen A und B.
H_o : Es besteht ein Zusammenhang zwischen A und B.

Für eine statistische Entscheidung benötigen wir nun die Wahrscheinlichkeit für den Zusammenhang zwischen A und B. Diese Wahrscheinlichkeit ist uns aber nicht bekannt; bekannt ist nur die Wahrscheinlichkeit für die Arbeitshypothese, die für die statistische Entscheidung ohne Bedeutung ist. Da wir die Wahrscheinlichkeit der Nullhypothese nicht aus der Wahrscheinlichkeit der Arbeitshypothese ableiten können, müssen wir uns behelfen.

Die Wahrscheinlichkeit für unsere Arbeitshypothese, die wir den Tabellen entnehmen können, rekurriert aus der Verteilung zweier unabhängiger Merkmale, deren Ausprägungen in einer Vielzahl von Fällen untersucht worden sind. Die Koeffizienten aus unserer Untersuchung sind nicht exakt gleich Null, sondern streuen in Form einer Normalverteilung um den Mittelwert Null herum. In den Tabellen können wir den Wert für r erfahren, bei

dem in 5% oder 1% der Fälle kein Zusammenhang mehr besteht. Transformieren wir den zugehörigen z-Wert, so können wir die Streuung dieser Verteilung mit dem Mittelwert Null ausrechnen. Innerhalb des Bereiches Mittelwert ± Streuung liegen dann 68% der Fälle, bei denen kein Zusammenhang besteht. Wir können nun, wie bei der Prüfung auf Normalverteilung, unsere Ergebnisse drei verschiedenen Klassen zuordnen:

Zuerst ist da die Klasse, in der der r-Wert eine Wahrscheinlichkeit kleiner 5% hat, bei der also ein Zusammenhang statistisch gesichert ist. Dann ist da die Klasse der r-Werte, die innerhalb des Bereiches ± 1 z, d.h. innerhalb der Streugrenzen liegen, zwischen denen sich 68% der Fälle von Unabhängigkeit befinden. Diese r-Werte können wir als im Sinne der Hypothese akzeptieren, d.h. diese r-Werte sollen die Nullhypothese widerlegen. Als dritte Klasse bleibt der Bereich zwischen der ersten und zweiten Klasse, der weder eine positive noch eine negative Entscheidung zuläßt.

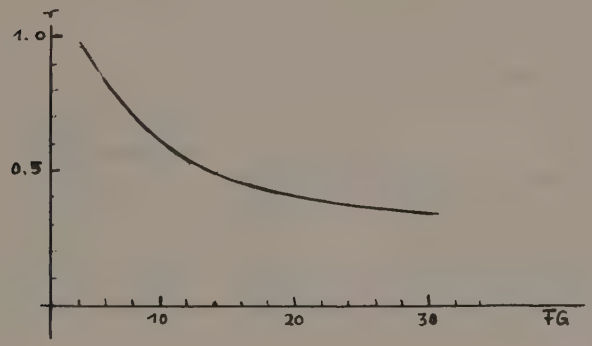
Die Grenzen zwischen den Klassen müssen wir selbst festlegen. Betrachtet man bei MITTENECKER [11] die Tafel VI auf Seite 19o, so sieht man, daß die Mindestgrößen von r, die signifikant sind (d.h. bei denen die Nullhypothese "kein Zusammenhang" eine Wahrscheinlichkeit von 5% oder 1% besitzt), von der Anzahl der Freiheitsgrade stark beeinflußt werden. In unserem Zusammenhang ergeben sich die Freiheitsgrade aus der Vpn-Zahl minus 1. Je größer die Anzahl der Vpn ist, desto kleiner muß der r-Wert sein, bei dem die Hypothese "kein Zusammenhang" eine Wahrscheinlichkeit von 5% oder 1% besitzt. Diese Anhängigkeit macht noch einmal deutlich, daß die statistische Entscheidung so angelegt wird, daß der ungünstige Fall wahrscheinlicher ist, d.h. daß man gegen den Zufall absichert. Bei wenigen Vpn kann rein zufällig ein Zusammenhang bestehen, der sich bei einer Untersuchung mit vielen Vpn nicht mehr herausstellt. Deshalb sind die kritischen Zahlen so gewählt, daß es bei wenigen Vpn fast unmöglich ist, einen Zusammenhang zu sichern, bei vielen jedoch nicht. Da wir in unserem Fall die Hypothesen

umgekehrt haben, ergibt sich daraus, daß wir unsere Arbeitshypothese bei wenigen Vpn nur selten widerlegen können, bei vielen jedoch häufig. Mit dieser Tatsache sichern wir uns aber nicht gegen den Zufall ab, sondern machen uns diesen zu Nutze. Diese Schwäche bei der Klasseneinteilung muß stets beachtet werden. Um haltbare Ergebnisse zu erzielen, müssen wir Versuche mit vielen Vpn heranziehen, damit uns nicht der Vorwurf gemacht werden kann, daß wir unter allen Umständen unsere Hypothese bestätigen wollen.

Für die Durchführbarkeit der hier vorgeschlagenen Methode spricht auf der anderen Seite, daß die Mindestwerte für r nicht linear abfallen. Bei sehr wenigen Vpn macht sich der Unterschied von 1 oder 2 Vpn schon stark bemerkbar; macht man Versuche mit vielen Vpn, so wird der Unterschied verschwindend gering. Die nachfolgende Abbildung gibt einen Überblick über die Abhängigkeit des r-Wertes von der Anzahl der Freiheitsgrade (=Zahl der Vpn minus 1)

Tabelle 12

Abhängigkeit der Mindestgrößen von r, die signifikant einen Zusammenhang andeuten, von der Anzahl der FG auf dem 5%-Niveau



Der Unterschied zwischen den r-Werten wird mit zunehmender Vpn-Zahl geringer. Der Unterschied zwischen 2 oder 11 Vpn, das sind 1 oder 1o FG, beträgt o.421, der zwischen 1o und 2o FG o.153 und der zwischen 2o und 3o FG nur noch o.o74. Von 2o FG an wird der Unterschied sehr gering, der durch eine Erhöhung der Vpn-Zahl erzielt wird. Bei unseren Versuchen schwanken die Vpn-Zahlen zwischen 15 und 25, sie reichen für eine vorläufige Analyse aus.

Entsprechend den verwendeten Vpn-Zahlen kann man für die einzelnen Werte die kritischen Zahlen berechnen. Die Grenze, ab der ein r-Wert signifikanten Zusammenhang anzeigt, wollen wir mit c_Z bezeichnen. Die Grenze, die durch die Streuung gesetzt wird, bezeichnen wir mit c_S. Den c_Z-Wert erhalten wir durch Nachschlagen in der Tabelle und den c_S-Wert können wir nach der Formel $z = (X - \mu) / \sigma$ berechnen.

Haben wir z.B. einen Versuch mit 2o Vpn, so ist der c_Z-Wert auf dem 5%-Niveau, das wir generell wählen wollen, r = o.433. In der obigen Formel entspricht dies dem X-Wert. Der zugehörige z-Wert ist 1.65 und der Mittelwert (μ) ist gleich Null. Setzen wir die Werte in die Formel ein, so erhalten wir für die Streuung den Wert o.262. Für die Entscheidung ergibt dies:
Alle r-Werte, die größer als +o.433 oder kleiner als -o.433 sind, zeigen Zusammenhang an;
alle r-Werte, die kleiner als +o.262 oder größer als -o.262 sind, sollen keinen Zusammenhang anzeigen;
alle r-Werte, die zwischen diesen Bereichen liegen, lassen keine Entscheidung zu (bei der Korrelationsrechnung ist der Zusammenhang umso größer, je mehr sich der r-Wert der Zahl 1 nähert, wobei +1 vollkommenen gleichgerichteten, -1 volkommenen entgegengesetzten Zusammenhang bedeutet; die kritischen Werte grenzen somit einmal gegen +1 ab, zum anderen aber auch gegen -1).

Für die bei den Versuchen von HRUSCHKA vorkommenden Vpn-Zahlen lassen sich folgende kritische Werte aufstellen:

Tabell 13

kritische Werte für die Vpn-Zahlen bei den Versuchen von HRUSCHKA

Vpn-Zahl	c_Z	c_S
15	o.497	o.3o1
16	o.482	o.292
18	o.456	o.276
25	o.388	o.235

Wir haben damit die statistischen Kriterien für die Entscheidung festgelegt. Es müssen nun die r-Werte für die einzelnen Versuche und die zu überprüfenden Größen berechnet werden. Das statistische Verfahren dafür ist die Produkt-Moment-Korrelation.

Bei der Berechnung selbst können wir auf die Umrechnungen, die bei der Prüfung auf Normalverteilung vorgenommen wurden, zurückgreifen. Für die relativen Bereichsumfänge benötigen wir die bezogenen Werte, für Systemumfang und Systemenden die unbezogenen. Die fraglichen Wertereihen werden zugeordnet und dann mit der Formel für den Produkt-Moment-Koeffizienten verrechnet.

Die Ergebnisse werden im Folgenden entsprechend der Einteilung durch die Hypothesen dargestellt:
(1) relativer Bereichsumfang - Systemumfang, (2) Systemanfang - Systemende, (3) Systemenden - relativer Bereichsumfang.

Tabelle 14

Korrelationskoeffizienten für die Korrelationen zwischen relativen Bereichsumfängen und Systemumfang

Versuch	Vpn	1.Kat.	2.Kat.	3.Kat.	4.Kat.	5.Kat.
Bleistifte						
1.	16	-o.216	-o.422	*+o.592*		
2.	15	(-o.119	+o.297	-o.1o1	-o.254)	
3.	25	-o.256	+o.339	-o.2o9	*+o.491*	-o.381
4.	15	-o.o81	+o.242	-o.251	*+o.56o*	-o.411
Taschentücher						
1.	15	+o.128	+o.o74	-o.o47	+o.o87	-o.o72
2.	15	+o.326	-o.413	+o.1oo		
3.	25	+o.228	-o.135	-o.o54		
4.	15	-o.344	+o.o43	+o.333		
Christbaumkugeln						
1.	15	-o.462	+o.455			
2.	18	+o.o18	+o.3oo	-o.432	-o.2o7	+o.198
3.	15	-o.299	-o.349	*+o.558*		
4.	25	-o.265	-o.164	*+o.4o2*		
Armbanduhren						
1.	15	+o.327	-o.321	+o.o48	+o.222	+o.o12
2.	25	-o.178	+o.o13	+o.136		
3.	25	+o.369	+o.o22	-o.292		
4.	15	-o.415	+o.118	+o.294		

Alle Werte, die oberhalb der kritischen Zahl c_Z liegen, also signifikant Zusammenhang anzeigen, wurden kursiv gedruckt.
Alle Werte, die unterstrichen sind, liegen innerhalb des durch c_S begrenzten Streubereiches.

2. Prüfung auf Unabhängigkeit der relativen Bereichsumfänge vom absoluten Systemumfang

In Tabelle 14 der vorangehenden Seite sind die Korrelationskoeffizienten im einzelnen aufgeführt. Entsprechend der Kategorienanzahl eines Versuches ergeben sich ebenso viele Koeffizienten, da die Korrelationen des Systemumfanges mit jedem Bereich zu berechnen waren. Die einzelnen r-Werte können nun entsprechend den zugehörigen kritischen Werten (siehe Seite 61) eingeteilt werden. Die Werte, die einen signifikanten Zusammenhang anzeigen, also größer als $+c_Z$ und kleiner als $-c_Z$ sind, wurden in der Tabelle durch kursiv gedruckte Zahlen gekennzeichnet. Die Werte, die innerhalb der Streugrenzen liegen, also kleiner als $+c_S$ und größer als $-c_S$ sind, wurden in der Tabelle unterstrichen. Die folgende Tabelle faßt die Ergebnisse noch einmal zusammen, indem sie die Häufigkeit anzeigt, mit der die einzelnen Klassen vertreten sind:

Tabelle 15

Häufigkeit der Korrelationskoeffizienten für die durch c_Z und c_S begrenzten Klassen (ausgehend von den Absolutwerten)

Klasse	f
oberhalb c_Z	5
unterhalb c_S	33
zwischen c_Z und c_S	20
Summe	58

Die Tabelle geht von den Absolutwerten aus, weil dadurch die Klasseneinteilung vereinfacht wird. Alle Zahlenwerte, die größer als c_Z

sind, zeigen signifikanten Zusammenhang an; damit sind die Werte gemeint, die im positiven Bereich größer im negativen Bereich jedoch wertmäßig kleiner sind.

Von den insgesamt 58 hier berechneten Koeffizienten zeigen nur 5 signifikanten Zusammenhang an, mehr als die Hälfte aber, nämlich 31, liegen innerhalb der Streugrenzen und zeigen entsprechend unserer Entscheidungskriterien Unabhängigkeit an.

Die Tendenz der Korrelationskoeffizienten kann außerdem noch an ihrem Mittel und der zugehörigen Streuung gezeigt werden. Bei vorhandener Unabhängigkeit sollte der Mittelwert o sein und die Streuung entsprechend der Vpn-Zahl dem Wert c_S entsprechen. Wir mitteln die Werte einmal getrennt für jede Kategorie und dann über alle Kategorien. Wenn wir die zugehörigen Vpn-Zahlen ebenfalls mitteln, können wir ungefähr die Streuung berechnen, die bei vorausgesetzter Unabhängigkeit vorhanden sein dürfte.
Die nachfolgende Tabelle zeigt die Mittelwerte und Streuungen für die einzelnen Kategorien getrennt.

Tabelle 16

Mittelwert und Streuung der Korrelationskoeffizienten, getrennt für jeden Bereich

	1. Kat.	2.Kat.	3.Kat.	4.Kat.	5.Kat.
$\bar{r}$	-o.o77	+o.o11	+o.o72	+o.o76	+o.132
σ	o.268	o.274	o.3o2	o.321	o.234
σ_E	o.276	o.276	o.269	o.283	o.276
N	16	16	15	6	5

σ_E : nach der gemittelten Vpn-Zahl zu erwartende Streuung, wenn Unabhängigkeit vorhanden.

Der mittlere Mittelwert über alle Kategorien beträgt -o.o1o, die zugehörige Streuung ist o.28o, die zu erwartende Streuung beträgt o.276.

Tabelle 17

Korrelationskoeffizienten für die Korrelation zwischen Systemanfang und Systemende

Versuch	Vpn	Systemanfang Systemende
Taschentücher		
1.	15	-o.482
2.	15	+o.1o1
3.	25	+o.o31
4.	15	-o.162
Christbaumkugeln		
1.	15	+o.o82
2.	18	(+o.224)
3.	15	(+o.132)
4.	25	(+o.151)
Armbanduhren		
1.	15	-o.328
2.	25	(+o.119)
3.	25	-o.29o
4.	15	*-o.742* bei 8 Herren *-o.763*
		bei 7 Damen (+o.188)

Alle Werte, die signifikanten Zusammenhang anzeigen, wurden kursiv gedruckt.
Alle Werte, die innerhalb der Streugrenzen liegen, wurden unterstrichen.

kritische Werte für den vierten Uhrenversuch

Vpn	c_S	c_Z
8	o.4o4	o.666
7	o.429	o.7o7

3. Prüfung auf Unabhängigkeit des Systemanfanges vom Systemende

In der vorangehenden Tabelle 17 sind die Einzelergebnisse aufgeführt und entsprechend den kritischen Werten gekennzeichnet. Bei dem 4. Uhrenversuch wurde das Gesamtergebnis noch nach Damen und Herren aufgeschlüsselt, um für die spätere Diskussion mehr Einblick in die zugrundeliegenden Verhältnisse zu erhalten. Die kritischen Werte für die durch die Aufschlüsselung verringerte Vpn-Zahl sind ebenfalls in Tabelle 17 angegeben.

Die Häufigkeit, mit der die Koeffizienten signifikanten Zusammenhang anzeigen, innerhalb des Streubereiches bleiben oder zwischen diesen Bereichen liegen, wird in Tabelle 18 dargestellt.

Tabelle 18

Häufigkeit der Korrelationskoeffizienten
für die durch c_Z und c_S begrenzten Klassen
(ausgehend von den Absolutwerten)

Klasse	f
oberhalb c_Z	1
unterhalb c_S	8
zwischen c_Z und c_S	3
Summe	12

Von den hier berechneten Koeffizienten zeigt einer Zusammenhang an und 8 von 12 liegen innerhalb der Streugrenzen und zeigen Unabhängigkeit an. Der Mittelwert aller Koeffizienten beträgt -o.o4o, die zugehörige Streuung o.225. Die zu erwartende Streuung bei vorausgesetzter Unabhängigkeit beträgt o.269.

Korrelationskoeffizienten für die Korrelation zwischen Systemanfang und relativer Bereichsgröße

Versuch	Vpn	1.Kat.	2.Kat.	3.Kat.	4.Kat.	5.Kat.
Taschentücher						
1.	15	+o.o16	-o.o22	-o.o55	-o.o66	+o.178
2.	15	-o.463	+o.321	+o.133		
3.	25	*-o.433*	+o.o83	+o.246		
4.	15	-o.137	+o.o1o	+o.141		
Christbaumkugeln						
1.	15	+o.493	-o.479			
2.	18	(+o.169	-o.236	+o.24o	+o.o75	-o.222)
3.	15	(+o.44o	+o.292	*-o.62o*)		
4.	25	(+o.223	+o.o78	-o.27o)		
Armbanduhren						
1.	15	-o.4o4	*+o.567*	+o.o52	+o.312	-o.329
2.	25	(+o.18o	+o.oo4	-o.15o)		
3.	25	+o.oo4	-o.o7o	+o.o7o		
4.	15	+o.23o	-o.o62	+o.216		
Bleistifte						
1.	16	+o.216	+o.422	*-o.592*		
2.	15	(+o.119	-o.297	+o.1o1	+o.254)	
3.	25	+o.231	-o.282	+o.171	*-o.426*	+o.313
4.	15	+o.o81	-o.242	+o.251	*-o.56o*	+o.411

Alle Werte, die signifikanten Zusammenhang anzeigen, wurden kursiv gedruckt.
Alle Werte, die innerhalb der Streugrenzen liegen, wurden unterstrichen.

Tabelle 2o

Korrelationskoeffizienten für die Korrelation zwischen Systemende und relativer Bereichsgröße

Versuch	Vpn	1.Kat.	2.Kat.	3.Kat.	4.Kat.	5.Kat.
Taschentücher						
1.	15	-o.193	+o.131	+o.143	-o.o17	-o.167
2.	15	+o.125	+o.3o6	-o,444		
3.	25	+o.3o4	+o.13o	-o.33o		
4.	15	+o.487	-o.o56	-o.478		
Christbaumkugeln						
1.	15	+o.224	-o.225			
2.	18	(-o.118	-o.243	+o.4o6	+o.218	-o.117)
3.	15	(+o.127	+o.263	-o.34o)		
4.	25	(+o.2o6	+o.154	-o.343)		
Armbanduhren						
1.	15	+o.1o2	-o.343	-o.142	-o.123	+o.451
2.	25	(+o.o38	-o.o63	+o.o1o)		
3.	25	*-o.514*	+o.11o	+o.255		
4.	15	+o.o71	-o.o25	+o.o27		

Alle Werte, die signifikanten Zusammenhang anzeigen, wurden kursiv gedruckt.
Alle Werte, die innerhalb der Streugrenzen liegen, wurden unterstrichen.

4. Prüfung auf Unabhängigkeit der Systemenden von der relativen Bereichsgröße

Tabelle 19 zeigt die Einzelergebnisse der Korrelationen zwischen Systemanfang und relativer Bereichsgröße, Tabelle 2o die Korrelationen zwischen Systemende und relativer Bereichsgröße. Die einzelnen Koeffizienten wurden entsprechend den kritischen Werten gekennzeichnet. Die folgende Tabelle 21 faßt die Ergebnisse zusammen, indem sie aufzeigt, wie häufig die einzelnen Klassen mit Koeffizienten besetzt sind.

Tabelle 21

Häufigkeit der Korrelationskoeffizienten
für die durch c_Z und c_S begrenzten Klassen
(ausgehend von den Absolutwerten)

Klasse	f (Systemanfang)	f (Systemende)
oberhalb c_Z	6	1
unterhalb c_S	37	28
zwischen c_S und c_Z	15	12
Summe	58	41

Die Gesamtzahl der Koeffizienten unterscheidet sich bei der Korrelation mit dem Systemanfang und der mit dem Systemende, da bei den Bleistiftversuchen nur ein Systemende vorhanden ist.

Einer bzw. sechs Werte zeigen Zusammenhang an, demgegenüber deuten aber jeweils mehr als die Hälfte der Werte Unabhängigkeit an.

Die Mittelwerte über die einzelnen Korrelationskoeffizienten, getrennt für jede Kategorie und für Systemanfang und -ende zusammen mit den zugehörigen Streuungen und den bei vorausgesetzter Unabhängigkeit zu erwartenden Streuungen, sind in der folgenden Tabelle zusammengestellt.

Tabelle 21

Mittelwert und Streuung der Korrelationskoeffizienten, getrennt für jeden Bereich

a. Korrelationen zwischen Systemanfang und relativen Bereichen

	1.Kat.	2.Kat.	3.Kat.	4.Kat.	5.Kat.
$\bar{r}$	+o.o5o	+o.oo5	-o.oo4	-o.o69	+o.o7o
σ	o.223	o.217	o.222	o.284	o.277
σ_E	o.276	o.276	o.269	o.283	o.276
N	16	16	15	6	5

b. Korrelationen zwischen Systemende und relativen Bereichen

	1.Kat.	2.Kat.	3.Kat.	4.Kat.	5.Kat.
$\bar{r}$	+o.o71	+o.o11	-o.112	+o.o26	+o.o56
σ	o.179	o.171	o.255	o.123	o.26o
σ_E	o.269	o.269	o.269	o.292	o.292
N	12	12	11	3	3

Der mittlere Mittelwert beträgt bei der Korrelation mit dem Systemanfang +o.o1o, die zugehörige Streuung ist o.245, die zu erwartende Streuung o.276.

Der mittlere Mittelwert bei der Korrelation mit dem Systemende beträgt +o.o1o, die zugehörige Streuung ist o.198, die zu erwartende Streuung ist o.276.

II. Diskussion der Ergebnisse

Betrachtet man die erzielten Ergebniss in ihrer Gesamtheit, so wurden 169 Korrelationskoeffizienten berechnet. Von diesen 169 Koeffizienten zeigen 7% (13) signifikanten Zusammenhang an, 62% (1o5) liegen innerhalb der Streugrenzen und 3o% (51) liegen zwischen den beiden Bereichen. Die Mehrzahl der Koeffizienten liegt also innerhalb der Streugrenzen und läßt die Tendenz der Nullkorrelation oder Unabhängigkeit erkennen.

Bei der Prüfung auf Normalverteilung stellte sich heraus, daß bei einer Reihe von Versuchen ein Systemende häufig nicht besetzt wurde und so ein fast konstanter Wert für dieses Ende herauskam (siehe Seite 47). Berechnet man nun mit diesem Ende eine Korrelation, so erhält man eine Scheinunabhängigkeit, weil die Korrelation einer variierenden Wertereihe mit einer Konstanten stets gleich Null ist. Diese Scheinunabhängigkeiten sind zwar im Sinne der Hypothese, aber nicht durch zugrundeliegende Bedingungsgefüge entstanden, sondern durch Besonderheiten der Versuchskonzeption. Für die weitere Betrachtung müssen diese Korrelationen herausgenommen werden, da sie keinen Aussagewert haben.

Es ergaben sich zwar nur sehr wenige Zusammenhang anzeigende Koeffizienten, aber gerade diese Korrelationen sollte man näher untersuchen, da sie für spätere Versuchsplanungen aufschlußreich sein können. Das Versuchsmaterial, an dem die Korrelationen berechnet wurden, ist nicht in Hinblick auf diese Berechnungen erstellt worden. Unregelmäßigkeiten der Versuchskonzeption können somit zu Zusammenhang anzeigenden Korrelationen führen. Diese Möglichkeit muß deshalb überprüft und abgesteckt werden, damit bei kommenden Versuchskonzeptionen darauf geachtet werden kann.

Im Folgenden werden einmal die Scheinkorrelationen herauszunehmen und dann die Zusammenhang anzeigenden Korrelationen zu analysieren sein.

1. Korrelationen mit einer Konstanten

Tabelle 12 (Seite 47) gibt an, wie oft die Systemenden nicht benutzt wurden. Es müssen nun für die hier zur Diskussion stehenden Untersuchungen die Ergebnisse herausgenommen werden, bei denen eher eine Korrelation mit einer Konstanten als mit einer frei variierenden Wertereihe vorliegt.

Das Systemende ist durchgängig verwendet worden, es stellt somit eine frei variierende Wertereihe dar.

Die Bleistiftversuche nehmen eine Sonderstellung ein, da bei ihnen sowieso nur ein Systemende vorhanden ist, das selbst häufig nicht benutzt wurde. Dadurch wird nicht nur die Korrelation des Systemendes mit den relativen Bereichsumfängen, sondern auch die des Systemumfanges mit denselben beeinflußt, da dieser direkt vom Systemende abhängig ist (siehe Seite 49). Herauszunehmen sind bei den Bleistiftversuchen deshalb auf jeden Fall alle Ergebnisse des zweiten Versuchs, weil hier bei insgesamt 15 Vpn 9mal das Systemende nicht benutzt wurde.

Bei den übrigen Versuchen sind folgende Ergebnisse herauszunehmen: Die Korrelationen des Systemanfanges mit dem Systemende und den relativen Bereichsumfängen des 2., 3. und 4. Versuchs mit den Christbaumkugeln (18 Vpn - 11mal nicht verwendet, 15 Vpn - 7mal nicht, 25 Vpn - 12mal nicht), des 2. Uhrenversuches und der Hälfte des vierten Uhrenversuches, bei dem Damenuhren beurteilt werden (25 Vpn - 9mal nicht, 7 Vpn - 4mal nicht).

Betrachten wir die übrigbleibenden Ergebnisse, so wird die Zahl der signifikanten Zusammenhang anzeigenden Koeffizienten absolut um 1 verringert, prozentual ausgedrückt aber auf 9% erhöht. Die Zahl der innerhalb der Streugrenzen liegenden Koeffizienten wird um 34 verringert und macht nur noch 55% der Fälle aus. Die Zahl der zwischen den Bereichen liegenden Koeffizienten wird um 5 verringert, ihr Anteil wird damit auf 36% erhöht. Die insgesamt 4o Koeffizienten, die herausgenommen werden mußten, wurden in den Tabellen bereits eingeklammert.

2. Zusammenhang anzeigende Koeffizienten

Die Untersuchung dieser Koeffizienten macht es notwendig, detailliert auf das zugrundeliegende Datenmaterial einzugehen. Der Übersichtlichkeit halber werden die Korrelationen entsprechend den Hypothesen getrennt betrachtet.

a. Korrelation zwischen relativen Bereichsgrößen und Systemumfang

Die in diesem Zusammenhang berechneten Korrelationen entstammen vier Versuchstypen, Versuchen mit Bleistiften, Taschentüchern, Christbaumkugeln und Armbanduhren. Die nachfolgende Tabelle schlüsselt die Koeffizienten entsprechend dem Versuchstyp auf:

Tabelle 22

Häufigkeit der Korrelationskoeffizienten für die durch c_Z und c_S begrenzten Klassen, getrennt nach Versuchstypus (ausgehend von den Absolutwerten)

Klasse		Bleistifte	Taschentücher	Christbaumkugeln	Armbanduhren
$r > c_Z$	f	3	o	2	o
	%	24	o	15	o
$r < c_S$	f	5	1o	5	9
	%	38	71	4o	64
$c_Z > r > c_S$	f	5	4	6	5
	%	38	29	45	36

Von der Verteilung der Koeffizienten in den einzelnen Klassen lassen sich die Versuchstypen folgendermaßen zusammenfassen:

Die Versuche mit den Bleistiften und Christbaumkugeln weisen in etwa gleich viele Zusammenhang anzeigende Koeffizienten (2 und 3) und gleich wenige im Streubereich liegende Koeffizienten auf.

Die Versuche mit den Taschentüchern und Armbanduhren erbringen beide keine Zusammenhang anzeigende Koeffizienten und in etwa gleich viele im Streubereich liegende Koeffizienten (9 und 1o).

Es ergibt sich also eine gewisse Unterscheidung zwischen den Versuchen mit den Bleistiften und Christbaumkugeln auf der einen Seite und denen mit den Taschentüchern und Armbanduhren auf der anderen Seite.

Schlägt man Tabelle 14 (Seite 62) auf, so kann man näher bestimmen, wo die Zusammenhang anzeigenden Koeffizienten liegen:

Bei den Bleistiftversuchen liegt ein Koeffizient bei der Korrelation von Systemumfang mit der 3. Kategorie, der Kleinkategorie (insgesamt 3 Kategorien), die beiden anderen liegen bei der Korrelation des Systemumfanges mit der 4. Kategorie, der Mittel-Klein-Kategorie (insgesamt 5 Kategorien). Die beiden nachfolgenden Korrelationen mit der 5. Kategorie liegen ebenfalls sehr hoch, nur eben unter der Signifikanzgrenze.

Bei den Versuchen mit den Christbaumkugeln ergeben sich beide Koeffizienten bei der Korrelation des Umfanges mit der 3. Kategorie, der Kleinkategorie (insgesamt 3 Kategorien). Alle fünf Zusammenhang anzeigende Koeffizienten sind positiv, die beiden ebenfalls hoch liegenden Koeffizienten der 5. Kategorie bei den Bleistiftversuchen sind negativ.

Die positive Korrelation besagt, daß ein großer Systemumfang einhergeht mit einem großen Bereichsumfang, ein kleiner Systemumfang mit einem kleinen Bereichsumfang. Da in allen fünf Fällen der diesbezügliche Bereichsumfang im Kleinbereich der Systeme zu finden ist, gilt es, die Frage abzuklären, wieso die Größe des Kleinbereiches einen positiven Zusammenhang mit der Größe des Systemumfanges aufweist.

Die Versuchsreihe von HRUSCHKA begann mit den Bleistiftversuchen. Dabei stieß man auf das Problem der Systembegrenzungen, als verschiedene Vpn sehr kleine Bleistifte nicht mehr mit einem kategorialen Urteil versehen wollten. Für diese Vpn waren die dargebotenen Gebilde keine Bleistifte mehr.

Es stellte sich daraufhin die Frage, wieso die eine Gruppe der Vpn alle Bleistifte beurteilte, die andere Gruppe aber nur die, die eine kritische Größe nicht unterschritten. Man analysierte die Gegebenheit 'Bleistift' und kam zu dem Schluß, daß man den dargebotenen Stift unter zwei Aspekten beurteilen konnte: einmal den Stift als Schreibwerkzeug, zum

anderen den Stift als Holzstückchen verschiedener Länge. In den großen Bereichen decken sich diese beiden Gesichtspunkte, erst im Kleinbereich kommt man zu Stiften, die als Schreibwerkzeug nicht mehr zu gebrauchen sind, die aber noch eine beurteilbare Holzstücklänge aufweisen. Wird der Bleistift unter dem Gesichtspunkt 'Schreibwerkzeug' beurteilt, so werden sehr kleine Stifte nicht mehr akzeptiert, der Kleinbereich hört eher auf und das ganze System ist kürzer, da im Großbereich kein beliebiger Anfangspunkt vorhanden ist, sondern alle Vpn mit dem Normbleistift beginnen. Wird demgegenüber das Holzstückchen beurteilt, so wird kein Stift abgelehnt, der Kleinbereich wird größer und damit auch der Umfang des gesamten Systems. Diese Abhängigkeit kann sich in einer positiven Korrelation zwischen Systemumfang und Bereichsgröße des Kleinbereiches ausdrücken.

Beruhen die Beurteilungen der einzelnen Vpn auf verschiedenen Beurteilungsaspekten, so ist es im Nachhinein kaum möglich festzustellen, welcher Aspekt an welcher Stelle wirksam wurde, und eine Interpretation der Ergebnisse wird äußerst schwierig. HRUSCHKA versuchte deshalb bei den folgenden Versuchen, die speziell im Hinblick auf das Problem der Systemenden angelegt wurden, durch die Versuchskonzeption nur einen einzigen Beurteilungsaspekt wirksam werden zu lassen. Untersucht werden sollten Gegebenheiten unterschiedlicher Form und Dimension, einmal das Quadrat und der Kreis und zum anderen die Kugel. Als Gegebenheiten wählte man dafür das Taschentuch (Quadrat), die Armbanduhr (Kreis) und die Christbaumkugel (Kugel). Bei den Vorversuchen stellte sich heraus, daß bei einer allgemein gehaltenen Versuchsanweisung wiederum mehrere Aspekte für die Beurteilung zum Tragen kamen. Die Versuchsanweisungen wurden daraufhin so formuliert, daß ein ganz spezieller Aspekt für die Beurteilung gefordert wurde, der dem persönlichen System der Vpn entsprach. Bei den Uhren und Tüchern erreichte man dies durch die geschlechtsspezifische Unterscheidung in Damen- und Herrenuhren oder -tücher. Bei den Christbaumkugeln setzte man als direkten Bezug den Weihnachtsbaum. Außerdem wies man ganz konkret auf die Möglichkeit nicht zum System gehören-

der Gegebenheiten hin.

Folgende Instruktionen wurden gegeben:

Bei den Taschentüchern: ... Ich wäre Ihnen dankbar, wenn Sie mir sagen würden, ob es (Tuch) für Sie, ganz speziell für Sie, ein großes, ein mittleres oder ein kleines Damen-(Herren) Taschentuch ist. Sollte eins dabei sein, das Ihnen als Damen-(Herren) Taschentuch ungeeignet erscheint, weil es zu klein oder zu groß ist, dann sagen Sie mir es bitte.....

Bei den Armbanduhren: ... Bitte, würden Sie mir sagen, ob die Uhr, die Sie sehen, Ihrer Meinung nach eine große, mittlere oder kleine Herren-(Damen) Armbanduhr ist.
... Wir meinen Armbanduhren, wie man sie für den täglichen Gebrauch trägt, regelrechte Armbanduhren, keine Taschenuhren und keine Ring- oder Fingeruhren, wie sie es als Schmuckstücke für Damen gibt. Sollte eine Uhr dabei sein, die Ihnen als Herren-(Damen) Armbanduhr ungeeignet erscheint, weil sie zu klein oder zu groß ist, dann sagen Sie es mir bitte.....

Bei den Christbaumkugeln: Ich lege jetzt hier auf das Tuch eine Christbaumkugel und möchte Sie bitten, mir zu sagen, ob diese Christbaumkugel groß, mittel oder klein ist. ... Dabei wollen wir nur Kugeln für Weihnachtsbäume, wie man sie im allgemeinen in Familien findet, beurteilen. Also gewöhnliche Weihnachtsbäume. Sollte mal eine Kugel dabei sein, von der Sie meinen, daß sie zu groß oder zu klein für einen solchen Baum ist, dann sagen Sie es mir bitte.....

Die drei Gegebenheiten unterscheiden sich von der Instruktion und somit vom Beurteilungsaspekt in einem Punkt ganz wesentlich. Bei den Tüchern und Uhren wird das ganz persönliche System angesprochen und in dem Satz,'Sollte eins dabei sein, das Ihnen ungeeignet erscheint', hervorgehoben. Bei den Kugeln ist dieser Satz jedoch dahingehend verändert wor-

den, daß hier auf einen allgemein üblichen Baum verwiesen wird: 'sollte eine dabei sein, von der Sie meinen, daß sie für einen solchen Baum ungeeignet ist,...'. Wäre das persönliche System angesprochen worden, so hätte man die Instruktion so formulieren müssen: 'sollte eine Kugel dabei sein, die Ihnen für Ihren Baum ungeeignet erscheint'. Jede Vp hat eine Weihnachtsbaumgröße, die für sie ideal ist. Entsprechend dieser Größe hat sie dasSystem der zugehörigen Kugeln ausgebildet. Ihr persönliches Bezugssystem 'Christbaumkugeln' ist abhängig von der Baumgröße, die sie bevorzugt. Wird nun auf den allgemein üblichen Weihnachtsbaum verwiesen, so muß die Vp bei ihrer Beurteilung alle möglichen Baumgrößen miteinbeziehen und davon das System der Kugelgrößen ableiten. Es wird somit nicht das persönliche, stabilisierte System angesprochen, sondern ein aus vielen Informationen abgeleitetes System.

Außerdem ist bei den Christbaumkugeln der Bereich der dargebotenen Kugeln so gewählt worden, daß viele große Kugeln durchgängig akzeptiert wurden. Wie bei den Bleistiften liegt hier somit quasi ein fester Ausgangspunkt für die Beurteilung im Großbereich vor. Die Vpn, die diese Versuche gemacht haben, kommen alle aus dem selben Kulturkreis, für sie müßte der allgemein übliche Weihnachtsbaum demnach auch in etwa die gleiche Größenvariation haben. Ausgehend von einem festen Anfangspunkt kann man sagen, erst im Kleinbereich 'scheiden sich die Geister', hier können eventuell vorhandene Unterschiede im persönlichen System wirksam werden. Bevorzugt eine Vp kleine Bäume und hat dementsprechend ein zum Kleinbereich hin verlagertes persönliches Bezugssystem, so sind ihr die kleinen Kugeln recht vertraut und sie wird sie kaum als nicht mehr zum System gehörend ablehnen. In Relation zum allgemein üblichen Baum sieht sie diese Kugeln aber als klein an und bringt mehr Klein-Urteile als andere Vpn, wodurch sich der Kleinbereich und der Gesamt-umfang vergrößern. Vpn, die demgegenüber große Bäume bevorzugen, sind mit den kleinen Kugeln wenig vertraut, sie lehnen sie eher ab, wodurch ihr Kleinbereich nicht vergrößert wird gegenüber den anderen Bereichen und der gesamte Umfang kleiner als bei anderen Vpn ist. Diese Abhängigkeit kann sich wiederum in einer positiven Korrelation zwischen Systemumfang und Bereichsumfang im Kleinbereich ausdrücken.

b. Korrelation zwischen Systemanfang und -ende

Bei der Korrelation zwischen den Systemenden fallen die Bleistiftversuche ganz heraus, da sie nur ein einziges Systemende besitzen. Bei den übrigen drei Gegebenheiten fallen viele Korrelationen aufgrund der Konstantenwirkung weg. Die Häufigkeitsverteilung entsprechend der Klassen der Koeffizienten sieht folgendermaßen aus:

Bei den Taschentüchern fällt kein Koeffizient weg, es liegen drei innnerhalb der Streugrenzen und einer im Bereich dazwischen.
Von den Christbaumkugeln kann nur ein Ergebnis beibehalten werden, das innerhalb der Streugrenzen liegt.
Die Armbanduhrenergebnisse sind um die Koeffizienten vermindert, die von den Versuchen mit Damenuhren herstammen. Die Koeffizienten der Herrenuhren liegen zweimal zwischen den beiden Bereichen und einer zeigt Zusammenhang an. Der Zahlenwert dieses Koeffizienten liegt sehr hoch, da für die Berechnung nur 8 Vpn herangezogen werkonnten. Er liegt jedoch mit o.763 nur geringfügig über dem kritischen Wert von o.7o7 und fügt sich somit in den Rahmen der übrigen Zusammenhang anzeigenden Koeffizienten.
Das Ergebnis des Christbaumkugelversuches kann man bei der weiteren Betrachtung außer Acht lassen, da bei dieser Gegebenheit nur ein Ergebnis vorliegt und nicht weiter interpretierbar ist.

Die Ergebnisse von Taschentuch- und Uhrenversuch unterscheiden sich. Die Uhrenkoeffizienten tendieren zu einer negativen Korrelation hin. Dies bedeutet: ein kleiner Systemanfang geht einher mit einem großen Systemende und umgekehrt.

Die Systeme sind demnach in ihrer Lage einmal mehr nach rechts hin verlagert, was einen kleinen Systemanfang und ein großes Systemende bewirkt, zum anderen mehr nach links verlagert, was einen großen Systemanfang und ein kleines Systemende bewirkt. Die Größe der Systeme selbst

muß dabei in etwa gleich sein. Betrachtet man die graphische Darstellung dieser Versuche, so fällt eine gewisse allgemeine Gleichheit in der Größe der einzelnen Systeme auf. Diese Gleichheit ist bei den Taschentuchversuchen nicht zu finden, auch nicht bei den herausgenommenen Damenuhrenversuchen. Da nun die Taschentücher, deren Koeffizienten wir berechneten, nur von Damen beurteilt wurden, könnte man auf die Idee kommen, daß hier ein geschlechtsspezifischer Unterschied vorliege. Die Herren haben in etwa gleich große Systeme ausgebildet, die unterschiedlich gelagert sind, die Damen nicht. Diese Annahme erscheint aber aufgrund der wenigen Daten und des noch nicht einmal durchgängig nachgewiesenen Zusammenhanges sehr weit hergeholt.

Eine andere Tatsache kommt hier vielleicht eher zur Geltung:
Die Versuche wurden in der frühen Nachkriegszeit durchgeführt. Während des Krieges wurden von Soldaten einheitliche, sehr große Militäruhren getragen. Vpn, die mit diesen Uhren viel Kontakt gehabt hatten, haben ein anderes Bezugssystem für Armbanduhren als solche, die diese Uhren kaum kannten. Da die Versuche kurz nach dem Kriege durchgeführt wurden, hatten die Vpn noch nicht ausreichend Gelegenheit, ihre Systeme wieder an die im zivilen Bereich üblichen Uhren anzupassen. Die Lage der Systeme im vorgegebenen Versuchsbereich bei Vpn mit großer Militäruhrenerfahrung mußte anders sein als die von Vpn, die im zivilen Bereich üblichen Uhrenkontakt hatten. Für die einen waren große Uhren das Normale, für die anderen im Vergleich dazu Kleine. Die unterschiedliche Lage wirkte sich auf die Systemenden aus, da einmal das System früh beginnt, der Systemanfang also klein ist, und früher aufhört, wodurch das Systemende groß wird, zum anderen aber das System erst später beginnt, der Anfang also groß wird, und später aufhört, wodurch das Ende klein wird. Dieser Effekt konnte deshalb zum Tragen kommen, weil Systeme im Zwischenbereich fehlten, bei denen eine mittlere Uhr das Normale darstellte.

Da man heute mehr über die Zusammensetzung der Vpn-Gruppen nicht mehr erfahren kann, was für die Klärung dieses Punktes unbedingt notwendig

erscheint, ist es angebracht, diese Versuche zu wiederholen. Man sollte als Vergleichsmöglichkeit auch Beurteilungen von Herrentüchern durchführen, um geschlechtsspezifische Unterschiede abstekken zu können.

c. Korrelation zwischen Systemenden und relativen Bereichsgrößen

Bei diesem Teil der Untersuchung fallen eine ganze Reihe von Ergebnissen heraus, da bei ihnen ein Konstanteneffekt vorliegt. Unbetroffen davon sind die Taschentuchversuche. Bei den Versuchen mit den Bleistiften entfällt der 2. Versuch, bei den Uhrenversuchen ebenfalls der 2. und bei den Versuchen mit den Christbaumkugeln außer dem ersten alle übrigen.

Der Bleistiftversuch wurde nur mit einem Systemende durchgeführt, wodurch sich eine direkte Abhängigkeit zwischen Systemumfang und dem einen Systemende ergibt. Die Ergebnisse dieser Korrelationen entsprechen in ihrem Zahlenwert auch fast denen der Korrelation zwischen Systemumfang und Bereichsgrößen (siehe Tabelle 14, Seite 62). Für die Ergebnisse selbst trifft deshalb das in Abschnitt a. bereits Gesagte zu (Seite 74 ff).

Nimmt man bei dieser Betrachtung die Bleistiftversuche heraus, so ergibt sich für die Häufigkeitsverteilung folgendes Bild:

Tabelle 23

Häufigkeit der Korrelationskoeffizienten für die durch c_Z und c_S begrenzten Klassen, getrennt nach Systemanfang und -ende

Klasse	Systemanfang	Systemende
$r > c_Z$	2	1
$r < c_S$	16	17
$c_Z > r > c_S$	9	9
N	27	27

Die Ergebnisse der Korrelation mit dem Systemanfang oder dem Systemende unterscheiden sich nicht. In der überwiedenden Mehrzahl der Fälle liegen die Koeffizienten innerhalb der Streugrenzen.

Die drei signifikanten Zusammenhang anzeigenden Koeffizienten treten einmal bei den Taschentuchversuchen und zweimal bei den Uhrenversuchen auf. Da bei den Christbaumkugeln nur noch ein Versuch übrig geblieben ist, kann es Zufall sein, daß hier kein Ergebnis Zusammenhang anzeigt.

Bei dem Taschentuchversuch ergibt sich eine negative Korrelation zwischen der Größe der 1. Kategorie, der Großkategorie, und dem Systemanfang. Fällt die Großkategorie relativ groß aus, bleibt der Systemanfang klein,und umgekehrt.

Bei den Uhren zeigt sich einmal ein Zusammenhang zwischen der zweiten Kategorie, mittel-groß, und dem Systemanfang, der diesmal positiv ist, zum anderen ein negativer Zusammenhang zwischen der ersten Kategorie, groß, und dem Systemende.

Eine einleuchtende Erklärung ist für diese Zusammenhänge nicht zu finden; die bereits früher aufgeworfenen Probleme werden auch hier einen Einfluß haben. Es ist aber nicht abzuklären, welche Tatsache an welcher Stelle ihren Einfluß geltend macht.

3. Mittelwert und Streuung

Für eine Art Trendanalyse wurden von den errechneten Koeffizienten bereits die Mittelwerte und Streuungen berechnet. Stellt man einmal die mittleren Mittelwerte, die in der Darstellung der Ergebnisse aufgezeigt wurden, zusammen, so ergibt sich folgende Tabelle:

Tabelle 24

mittlere Mittelwerte und Streuungen der Korrelationskoeffizienten

Korrelation	μ	σ	σ_E	N
Systemumfang mit relativem Bereich	-o.o1o	o.28o	o.276	58
Systemanfang mit Systemende	-o.o4o	o.225	o.269	12
Systemanfang mit relativem Bereich	+o.o1o	o.245	o.276	58
Systemende mit relativem Bereich	+o.o1o	o.198	o.276	41

Die Mittelwerte liegen recht nahe an Null und die Streuungen durchweg unterhalb der Streuung, die bei vorhandener Unabhängigkeit zu erwarten ist. Die Tendenz scheint für Unabhängigkeit zu sprechen.

Es wurde aber bereits darauf hingewiesen, daß viele Koeffizienten nicht mitberücksichtigt werden dürfen, da bei ihnen ein Konstanteneffekt vorhanden ist (siehe Seite 73). Dieser Effekt muß sich natürlich auch auf die Mittelwerte auswirken, weil zu viele nullähnliche Koeffizienten dabei berücksichtigt wurden. Es ist nun interessant, die mittleren Mittelwerte ohne diese Koeffizienten zu berechnen. Die folgende Tabelle gibt die so erhaltenen Werte wieder.

Tabelle 25

mittlere Mittelwerte und Streuungen der Korrelationskoeffizienten ohne die Koeffizienten, bei denen ein Konstanteneffekt auftritt

Korrelation	μ	σ	σ_E	N
Systemumfang mit relativem Bereich	+0.022	0.253	0.276	54
Systemanfang mit Systemende	-0.224	0.237	0.276	8
Systemanfang mit relativem Bereich	+0.014	0.243	0.276	40
Systemende mit relativem Bereich	-0.007	0.218	0.283	27

Eine sehr deutliche Veränderung ist bei dem Mittelwert der Korrelation zwischen Systemanfang und -ende eingetreten. Bei dieser Wertereihe sind die positiven Koeffizienten weggefallen und die negativen überwiegen jetzt stark. Da das Gesamt-N sehr niedrig liegt, kann dieses Übergewicht an negativen Koeffizienten nicht ausgeglichen werden.

Bei den übrigen Korrelationen sind die Veränderungen nur sehr geringfügig, die Tendenz zur Nullkorrelation ist weiterhin erkennbar.

III. Zusammenfassung der Ergebnisse und Folgerungen für spätere Untersuchungen

Die vorangehende Analyse beschäftigte sich mit der Prüfung auf Unabhängigkeit zweier Größen voneinander, die in den folgenden Hypothesen postuliert wurde: 3. Jeder relative Bereichsumfang ist unabhängig vom absoluten Systemumfang.

4. Systemanfang und -ende sind unabhängig von einander.
5. Systemanfang und -ende sind unabhängig von den relativen Bereichsumfängen.

Ein Maß für die Abhängigkeit zweier Größen erhält man mit Hilfe der Korrelationsrechnung in Form eines Korrelationskoeffizienten. Die Anwendung der Korrelationsrechnung bringt jedoch in diesem Zusammenhang von der statistischen Entscheidung her Schwierigkeiten mit sich. Die Korrelationsrechnung ist darauf ausgerichtet, die Vermutung eines eventuell vorhandenen Zusammenhanges zu überprüfen. Sie geht dabei von dem der Vermutung konträren Fall des Nicht-Zusammenhanges aus und läßt die Hypothese 'Zusammenhang' nur dann zu, wenn die Wahrscheinlichkeit für 'Nicht-Zusammenhang' sehr klein ist. In unserem Fall geht die Vermutung dahin, daß kein Zusammenhang besteht, und für eine statistisch exakte Entscheidung muß die Wahrscheinlichkeit für Zusammenhang bekannt sein. Ist die Wahrscheinlichkeit dafür kleiner als ein vorher angesetzter kritischer Wert, so kann die Vermutung oder Hypothese 'kein Zusammenhang' als bestätigt gelten. Die Wahrscheinlichkeit für Zusammenhang war jedoch nicht bekannt und eine statistisch exakte Entscheidung konnte nicht erreicht werden. Ausgehend von der bekannten Wahrscheinlichkeit des 'Nicht-Zusammenhanges' wurden behelfsmäßig kritische Werte gesetzt, die wenigstens einen Überblick über die vorliegenden Verhältnisse gestatten sollen. Der erste kritische Wert, mit c_Z bezeichnet, war der Wert, der gesicherten Zusammenhang anzeigte. Der zweite kritische Wert, mit c_S bezeichnet, war die Streuung der bekann-

ten Verteilung des 'Nicht-Zusammenhanges', die aufgrund des c_S-Wertes berechnet werden konnte.

Die Größenpaare, die auf Unabhängigkeit überprüft werden sollten, erforderten alle Versuche mit Systemenden, wodurch wieder auf das Versuchsmaterial von HRUSCHKA allein zurückzugreifen war.

Als Ergebnis der Analyse lassen sich folgende Punkte herausheben:

1. Die Mittelwerte der Korrelationskoeffizienten legen eine Tendenz zur Nullkorrelation (= Unabhängigkeit) nahe.

2. Nur relativ wenige Koeffizienten zeigten gesicherten Zusammenhang an. Die meisten lagen innerhalb der durch $\pm$ 1δ gekennzeichneten Streugrenzen der Nullkorrelationsverteilung.

3. Die Tatsache, daß bei mehreren Versuchen ein Systemende häufig nicht verwendet wurde, wirkte sich bei den Korrelationskoeffizienten in Form eines Konstanteneffektes aus, der Null nahe Korrelationen erzeugt. Die fraglichen Koeffizienten wurden aus der Untersuchung ausgeklammert.

Für die Planung späterer Untersuchungen sollte man Folgendes beachten:

1. Für die statistische Entscheidung

Es wäre schön, wenn man wie für die Verteilung der Nullkorrelation ebenso für die Verteilung der perfekten Korrelation die zugehörige Normalverteilung ermittel könnte. Damit hätte man die Wahrscheinlichkeit der Nullhypothese und könnte dementsprechend exakte Entscheidungen auf dem 5%- oder 1%-Niveau fällen.

Da man diese Verteilung jedoch wohl kaum in absehbarer Zeit ermitteln wird, muß man bei der Entscheidung den Zusammenhang zwischen der Vpn-Zahl und dem Zusammenhang anzeigenden kritischen Werten be-

achten. Ist die Zahl der Vpn sehr klein, so liegt dieser Wert hoch, um die Hypothese 'Zusammenhang' gegen rein zufällig auftretende Zusammenhang anzeigende Koeffizienten abzusichern. Diese Absicherung gegen den Zufall wirkt sich aber bei der Umkehrung der Hypothese dahingehend aus, daß es mit einer kleinen Vpn-Zahl immer schwieriger wird, den für die Hypothese ungünstigen Fall zu erreichen. Ab einer gewissen Vpn-Zahl fällt dieser kritische Wert nur noch langsam. Diese Vpn-Zahl (3o oder 4o) sollte bei den Versuchen dann wenigstens erreicht werden.

2. Systemenden

Auch bei der Prüfung auf Unabhängigkeit ist wiederum unbedingt mit Versuchen mit echten Systembegrenzungen zu arbeiten, da sonst der Effekt einer Konstantenwirkung unrealistisch positive Ergebnisse erzielen läßt (die Korrelation mit einer Konstanten ist stets gleich Null).

3. Beurteilungsgesichtspunkte

Des weiteren sollte darauf geachtet werden, daß bei der Beurteilung nicht mehrere Aspekte zur Wirkung kommen können. Die Homogenität der Beurteilungsgrundlage bei den einzelnen Vpn ist unbedingt notwendig, da sich sonst unkontrollierbare Bedingungsverhältnisse bemerkbar machen können.

Die Gefahr, unrichtige Ergebnisse aufgrund verschiedener Beurteilungsgesichtspunkte zu erhalten, verringert sich jedoch beträchtlich, wenn echte Systembegrenzungen vorhanden sind. Erst wenn hierdurch die Stabilität der einzelnen Größen bereits beeinträchtigt worden ist, können andere Einflüsse richtig zur Geltung kommen.

C. Der Streuungsvergleich

I. Die statistische Bearbeitung und Diskussion der Befunde

Die Hypothese, die eine Aussage über die Streuung macht, lautet folgendermaßen:

6. Die intraindividuelle Streuung ist überaus signifikant kleiner als die interindividuelle Streuung.

Wird eine Gegebenheit von mehreren Vpn beurteilt, so werden die Unterschiede bei der Beurteilung mit interindividueller Streuung bezeichnet; es handelt sich um die Streuung, die von Individuum zu Individuum auftritt. Wird eine Gegebenheit von einer einzelnen Vp mehrmals beurteilt, so werden die dabei auftretenden Beurteilungsunterschiede mit intraindividueller Streuung bezeichnet; es handelt sich um die Streuung, die bei einem Individuum auftritt.

Die Hypothese besagt nun, daß die bei einzelnen Individuen auftretenden Streuungen, wenn sie eine Gegebenheit mehrmals beurteilen, wesentlich geringer sind als die Streuung, die von Individuum zu Individuum auftritt, wenn diese eine Gegebenheit zu beurteilen haben.

Die intraindividuelle Streuung kann man nur dann ermitteln, wenn die einzelnen Vpn eine Gegebenheit wenigstens zweimal beurteilen. Dies kann der Fall sein, wenn man einen Versuch unter gleichbleibenden Bedingungen wiederholt oder die einzelnen Versuchsgegebenheiten mehrmals darbietet.

Bei dem vorhandenen Datenmaterial wurde bei einem Versuch jede Gegebenheit einmal gezeigt, es kann hier also keine intraindividuelle Abweichung bestehen.

Tabelle 26

Übersicht über die Versuchsunterschiede bei den Versuchen von HRUSCHKA und BRÄUER mit Versuchswiederholungen

Nr.	Versuch	I	II	III	IV	V	Gegebenheit
HRUSCHKA							
1.	S. 49	2	4	15	54	E	Bleistifte
2.	S. 6o	5	5	15	54	E	Bleistifte
3.	S.132	2	3	15	71	E	Taschentücher
4.	S.137	2	3	15	37	E	Armbanduhren
5.	S.142	2	3	15	41	E	Christbaumkugeln
BRÄUER							
6.	11 + 14, 14J.	2	3	15	54	14	Bleistifte
7.	11 + 14, 14J.	2	5	1o	44	14	Bleistifte
8.	11 + 14, 11J.	2	3	15	54	11	Bleistifte
9.	11 + 14, 11J.	2	5	1o	44	11	Bleistifte
1o.	9 - 11, 9J.	2	5	2o	54	9	Bleistifte
11.	9 - 11, 1oJ.	2	5	2o	54	1o	Bleistifte
12.	9 - 11, 11J.	2	5	2o	54	11	Bleistifte
13.	12 - 14, 12J.	2	5	17	54	12	Bleistifte
14.	12 - 14, 13J.	2	5	2o	54	13	Bleistifte
15.	12 - 14, 14J.	2	5	16	54	14	Bleistifte
16.	Erwachsene	2	5	18	54	E	Bleistifte
17.	Klavier, 11J.	2	5	1o	29	11	Klaviertöne
18.	Klavier, 11J.	2	3	1o	29	11	Klaviertöne
19.	Klavier, 14J.	2	5	1o	29	14	Klaviertöne
2o.	Klavier, 14J.	2	3	1o	29	14	Klaviertöne

Legende:

I: Anzahl der Versuche unter gleichbleibenden Bedingungen

II: Anzahl der Kategorien bei diesem Versuch

III: Anzahl der Vpn

IV: Anzahl der dargebotenen Gegebenheiten pro Versuch

V: Alter der Vpn

E: Erwachsene

Für die Streuungsanalyse müssen die Versuche herangezogen werden, bei denen Wiederholungen durchgeführt wurden, also die selbe Gruppe von Vpn einen Versuch wenigstens zweimal beurteilte.

Durch diese Voraussetzung schrumpft das vorhandene Datenmaterial zusammen. Lediglich HRUSCHKA hat fünf Versuche und BRÄUER insgesamt 15 Versuche mit Wiederholungen durchgeführt. Die Tabelle 26, Seite 9o gibt einen Überblick über das Versuchsmaterial von HRUSCHKA und BRÄUER, auf die dort eingeführte Numerierung wird im Folgenden zurückgegriffen.

1. Einleitende Überlegungen

Für die Untersuchung dieser Hypothese müssen wir zwei Streuwerte miteinander vergleichen. Pro Versuch erhalten wir jeweils einen intra- und einen interindividuellen Wert, die paarweise zugeordnet sind. Das statistische Verfahren ist demnach der Vergleich von Paardifferenzen, der mit dem Wilcoxon-Test durchgeführt werden kann. Die Schwierigkeiten liegen bei der Überprüfung dieser Hypothese nicht bei den statistischen Verfahren selbst, sondern in der Methode, wie man zu den Streuwerten gelangt. Als Grundmaterial liegen uns verbale Urteile vor, deren Aussage wir numerisch erfassen müssen. Bei z.B. drei vorgegebenen Kategorien 'groß', 'mittel' und 'klein' vereinigt eine bestimmte Gegebenheit eine Anzahl von 'groß'-Urteilen und 'mittel'-Urteilen und vielleicht auch 'klein'-Urteilen auf sich. Die Schwierigkeit liegt nun darin, aus dieser Information ein Maß für die Streuung zu erhalten.

Da es verschiedene Möglichkeiten gibt, zu diesem Maß für die Streuung zu kommen, wird im Folgenden nicht mehr zwischen statistischer Bearbeitung und Diskussion getrennt. Die einzelnen Möglichkeiten werden vielmehr eingangs beschrieben, danach werden die damit erzielten Befunde dargestellt und anschließend werden diese Befunde diskutiert,

einmal von ihrem Aussagewert für die Hypothese her und zum anderen im Hinblick auf die zugrundeliegende Methode ihrer Gewinnung. Bei methodischen Schwächen ist es dann unter Umständen notwendig, die Methode zu modifizieren oder eine neue zu wählen. Da die Streuwerte, wie sich zeigen wird, nicht unabhängig von der Methode ihrer Gewinnung sind, ist eine strenge Trennung in Darstellung und Diskussion nicht möglich.

2. Die Methode von BRÄUER

Wir greifen zum großen Teil auf die Daten von BRÄUER zurück, weil dieser durchgängig Versuchswiederholungen durchgeführt hat. Diese Wiederholungen waren für seine Arbeit notwendig, da sie einen ähnlich gelagerten Aspekt wie den hier zur Diskussion stehenden untersuchen wollte. BRÄUER hat außerdem in seiner Arbeit selbst die Unterschiede bei den einzelnen Versuchsdurchgängen in der Beurteilung numerisch erfaßt und ist zu Maßen gelangt, die wir als Streuwerte benutzen können. Wir beginnen deshalb mit seiner Methode und werden sehen, in wie weit sie für die Beantwortung dieser Hypothese herangezogen werden kann. Zum besseren Verständnis ist es notwendig, auf die Grundgedanken seiner Arbeit näher einzugehen.

a. BRÄUER: Die Entwicklung von Bezugssystemen von der späten Kindheit an über Vorpubertät und Pubertät zum Erwachsenenalter.

Der Grundgedanke der BRÄUERschen Arbeit zielt auf die Entwicklung der Bezugssysteme ab. Die Bezugssysteme sind nicht von Kindheit an vorhanden, sondern sie entstehen im Laufe des Heranwachsens. Dieser Entwicklungsprozeß deutet sich einmal in der unterschiedlichen Festigkeit der Systeme an. Erwachsene haben stabile Systeme, Kinder eher instabile, die sehr anfällig gegen äußere und innere Einflüsse sind. Die Diskrepanz der einzelnen Systeme soll außerdem im Laufe ihrer Stabilisierung abnehmen, d.h. bei Kindern unterscheiden sich die persönlichen Systeme stark voneinander, bei den Erwachsenen liegen sie mehr auf einer Ebene.

Diese beiden Aspekte der Bezugssysteme, Stabilität des einzelnen und Diskrepanz untereinander, bezeichnet BRÄUER mit altersgruppentypischem und individualtypischem Stabilitätsfaktor. Der individualtypische Stabilitätsfaktor ist die Stabilität des Systemes einer Vp selbst, der altersgruppentypische Faktor ist Ausdruck für die Unterschiede der Systeme verschiedener Vpn einer Altersgruppe.

BRÄUER geht davon aus, daß sich beide Stabilitätsfaktoren im Laufe der Entwicklung von anfänglicher Instabilität oder Labilität zu Stabilität hin entwickeln.

Diesen Entwicklungsprozeß verfolgt er nun an Kindern und Heranwachsenden im Alter von 9 bis 14 Jahren und stellt diesen Erwachsene gegenüber. Er will so zu einer Information über die gruppentypische Ausprägung und Entwicklung der psychischen Dimension 'Kategorisierungsfähigkeit' und über persönlichkeitsspezifische Ausprägung und Entwicklung derselben gelangen.

Als Maß für den altersgruppentypischen Stabilitätsfaktor nimmt er die Beurteilungsunterschiede dieser Altersgruppe und bezeichnet sie mit interindividueller Labilität. Dem stellt er die intraindividuelle Labilität gegenüber, die sich aus individuellen Beurteilungsunterschieden dieser Altersgruppe herleitet. Bezeichnet man die Beurteilungsunterschiede mit Streuung, so ermittelt er einmal die interindividuelle Streuung einer Altersgruppe, zum anderen die intraindividuelle Streuung dieser Altersgruppe, womit die Verbindung zur sechsten Hypothese von WITTE vorhanden ist. BRÄUER ging zwar von einer anderen Intention aus, er wollte die altersbedingte Abnahme der Labilitäten bei inter- und intraindividueller Streuung aufzeigen, hat aber in seiner Arbeit bereits Maße für die beiden Streuungen ermittelt. Aus diesem Grunde mußte er auch bei all seinen Versuchen mit Wiederholungen arbeiten, da er sonst keine intraindividuellen Werte berechnen konnte. Das Ergebnis seiner Arbeit ist, daß die Stabilität sowohl inter- als auch

intraindividuell im Laufe der Entwicklung zunimmt und bei den Erwachsenen einen Höchstpunkt erreicht. Er ermittelte noch einige altersbedingte Besonderheiten bei dieser Entwicklung, grob gesehen steigt sie jedoch kontinuierlich an.

b. BRÄUERs Vorgehensweise

BRÄUER hat bereits die Werte für die beiden Streuungen ermittelt und wir können direkt auf diese Werte zurückgreifen und sie mit dem Wilcoxon-Test überprüfen. Es ist aber notwendig, die Methode ihrer Gewinnung eingehend darzustellen, da sie eventuell Schwächen aufweisen kann, die methodischer und inhaltlicher Natur sind.

Die Methode von BRÄUER erscheint recht plausibel und ist gut praktizierbar. Nehmen wir als Beispiel einen seiner Vorversuche mit neunjährigen Vpn:

Es handelt sich dabei um die Kategorisierung von Bleistiften. Insgesamt variierte der dargebotene Bereich über 52 Stifte, die fünf Kategorien zugeordnet werden sollten, den Kategorien 'ganz klein' (gkl), 'klein' (kl), 'mittel' (m), 'groß' (gr) und 'ganz groß' (ggr). Er zählt für jeden Bleistift getrennt aus, wie oft die einzelnen Kategorien diesem Stift zugeordnet wurden. Die Bleistifte sind numeriert, der größte ist Nummer 1 und der kleinste Nummer 52. Die Tabelle 27 auf der folgenden Seite zeigt das Ergebnis dieses Versuches. Die Kategorien sind wie oben (in Klammern) abgekürzt und die Summe der auf alle Kategorien fallenden Urteile muß stets 1o sein, da dieser Versuch mit 1o Vpn gemacht wurde. Da jede Vp 52 Stifte beurteilte, ergeben sich insgesamt 52o Beurteilungen.

Bleistift 1 z.B. wurde von allen Vpn mit 'ganz groß' beurteilt, Blei-

Tabelle 27

BRÄUER: Ergebnisse der Vorversuche, 9-jährige Vpn, 1o Vpn
Tabelle 1, Seite 14 bei BRÄUER

Bl.	gkl	kl	m	gr	ggr	Bl.	gkl	kl	m	gr	ggr
1					1o	28		2	8		
2				1	9	29		5	5		
3				2	8	3o		6	4		
4					1o	31		7	3		
5				1	9	32		6	4		
6				2	8	33		6	4		
7				5	5	34		7	3		
8				4	6	35		6	4		
9				8	2	36		9	1		
1o				7	3	37		9	1		
11			1	6	3	38		1o			
12			2	7	1	39		9	1		
13			3	7		4o	2	8			
14			2	5	3	41	2	8			
15			2	7	1	42	2	8			
16			3	7		43	7	3			
17			2	7	1	44	3	7			
18				8	2	45	6	4			
19			7	3		46	7	3			
2o			8	2		47	8	2			
21			7	3		48	9	1			
22			6	2	2	49	9	1			
23			1o			5o	1o				
24			9	1		51	1o				
25			8	2		52	1o				
26		1	7	2							
27		3	6	1							

N = 52o

stift 2 von 9 Vpn mit 'ganz groß' und von einer mit 'groß, usw.

BRÄUER erläutert seine Methode folgendermaßen:

"Wenn man davon ausgeht, daß es im Alltag der Absolutkategoriesierung nicht ausbleibt, daß ein Beurteiler eine Gegebenheit der Mittelkategorie, ein anderer sie aber der Nachbarkategorie 'groß' zuweist, daß sich diese beiden Menschen aber doch leicht bei Bemühung um eine feiner differenzierende Kategorisierung verständigen können, etwa derart, daß jener zugesteht, er meine eben 'reichlich mittel', und dieser, er meine eben 'gerade noch groß', dann wird man durch solche Vorkommnisse die interindividuelle Stabilität eines Systems nicht beeinträchtigt sehen. Mindere Stabilität wird man hingegen wohl annehmen, wenn einer Gegebenheit drei oder gar noch mehr Kategorien zugeordnet werden. Ja, es scheint nahezuliegen, die Anzahl der Beurteilungen eines Reizes, die auf 3 und mehr Kategorien entfallen, als Anhaltspunkt für das Ausmaß der Labilität des Systems innerhalb der Altersgruppe zu nehmen. Durch folgende Überlegung soll eine Quantifizierung der Labilität versucht werden.

Wenn wir eine Streuung der Urteile über zwei Kategorien als normal gelten lassen, dann wären die auf die 3. Kategorie entfallenden Urteile Hinweise auf Labilität. Beim Betrachten der Tabelle 1 (bei uns Tabelle 27, Seite 95) finden wir in Zeile 11 die erste Urteilsstreuung über 3 Kategorien. Nun ergibt sich die Frage, ob die linke oder rechte Abweichung (gesehen von der 'herrschenden' Kategorie 'groß') als Hinweis auf Labilität angesehen werden soll. Wir wollen annehmen, daß die Kategorie, die von den beiden abweichenden die meisten Urteile bekommt, von den meisten Beurteilern als 'richtig' betrachtet wird. Dann wäre also bei Bleistift 11 die Häufigkeit 1 in der Kategorie 'mittel' ein 'Außenseiter'.

M.a.W.: Als 'Außenseiter' (Hinweis auf Labilität) werden bei 3 Kategorien die geringsten Häufigkeiten gewertet, dazu alle eventuell auf die 4. und 5. Kategorie entfallenden Urteile.

Wenn man in Tabelle 1 die so umschriebenen Urteilshäufigkeiten zusammenfaßt, ergibt sich die Anzahl von 1o Urteilen, die als 'Außenseiter' anzusehen sind (B. 11,12,14,15,17,22,26,27).

Um nun dieses Maß für Labilität für alle Versuchsgruppen vergleichbar zu machen, wird der Prozentanteil an der Urteilsgesamtheit dieser Gruppe errechnet:

1o/52o = 1,9% (1o/52o) x 1oo = 1,9%

Die Labilität der interindividuellen Beurteilungen dieser Altersgruppe beträgt also 1,9%. Die interindividuelle Stabilität wäre der Ergänzungswert auf 1oo% und betrüge in dieser Altersgruppe 98,1%.

Neben diesem interindividuellen Stabilitätsmaß könnte in den Hauptversuchen noch ein intraindividuelles Maß entwickelt werden.

Wenn jede Versuchsperson jeden Bleistift zweimal beurteilt, dann werden vermutlich die zweiten Urteile über einen bestimmten Stift nicht immer auf die gleiche Kategorie entfallen wie die ersten Urteile. Der Anteil der Abweichungen zwischen 1. und 2. Darbietung eines Bleistiftes läßt sich als Maß für die intraindividuelle Sicherheit in der Zuordnung eines Reizes zu einer Kategorie interpretieren." (BRÄUER[5]), Seite 16 - 17)

Überträgt man die eben aufgezeichnete Vorgehensweise auf den Streuungsvergleich, so soll jede Abweichung um mehr als eine Kategorie als ein Punkt für Streuung gerechnet werden. Das Streuungsmaß ist so identisch mit dem Maß für Labilität. Die Hypothese von WITTE lautet in den Ausdrücken der BRÄUERschen Arbeit folgendermaßen:

> Die intraindividuelle Labilität ist überaus kleiner als die interindividuelle Labilität.

Die von BRÄUER erzielten Labilitätswerte sind in Tabelle 28 der folgenden Seite zusammengestellt.

Tabelle 28

Labilitätswerte der BRÄUERschen Versuche

Versuchs-Nr.	inter indiv. Labilität	intra indiv. Labilität	Anzahl der Kategorien	Alter
6	o,2%	o,o%	3	14
7	5,8%	2,8%	5	14
8	o,6%	o,o%	3	11
9	5,9%	5,5%	5	11
1o	1o,5%	4,4%	5	9
11	12,4%	6,3%	5	1o
12	5,8%	2,5%	5	11
13	5,8%	1,8%	5	12
14	7,6%	4,2%	5	13
15	5,9%	2,9%	5	14
16	3,1%	o,2%	5	Erwachsene
17	1o,4%	8,o%	5	11
18	2,6%	o,o%	3	11
19	7,2%	4,5%	5	14
2o	o,7%	o,o%	3	14

Ein methodischer Vergleich auf der Basis von Paardifferenzen erübrigt sich bei den gefundenen Labilitätswerten, da durchgängig die intraindividuelle Labilität kleiner als die zugehörige interindividuelle ist. Jeder Test kann somit nur 1oo %igen Unterschied anzeigen.

Die Größenordnung der Werte variiert stark bei den einzelnen Versuchen und Altersgruppen. Die interindividuellen Werte erstrecken sich dabei über einen größeren Spielraum als die intraindividuellen.

Bei Versuchen mit 3 Kategorien gibt es durchgängig keine intraindividuelle Abweichung über mehr als eine Kategorie.

Die BRÄUERschen Werte scheinen die Hypothese mehr als befriedigend zu bestätigen. Es ist jedoch verdächtig, daß nicht eine einzige Ausnahme besteht und es kommen Zweifel an der Methode selbst auf.

Bei der Erläuterung seiner Methode geht BRÄUER von einem Beispiel aus, bei dem keine Versuchswiederholung gemacht wurde. In seinen Hauptversuchen wird jeder Versuch einmal wiederholt und bei der Auszählung dieser Ergebnisse hat er einen methodischen Fehler gemacht, der im Zusammenhang seiner Arbeit keine große Rolle spielt, der aber für den Streuungsvergleich von grundlegender Bedeutung ist.

Wird ein Versuch mit 54 Bleistiften und 1o Vpn durchgeführt, so hat man pro Versuch 54o Urteile. Wiederholt man diesen Versuch, so hat man insgesamt 1o8o Urteile. Für die Berechnung des interindividuellen Labilitätsmaßes hat nun BRÄUER beide Versuche zusammengefaßt, er hat die Abweichungen innerhalb dieser 1o8o Urteile ausgezählt. Die Quersumme über die einzelnen Kategorien bei jeder Gegebenheit war bei 1o Vpn 2o, da beide Urteile der Vpn zusammengefaßt wurden. Hat nun eine Vp eine Gegebenheit in beiden Durchgängen stark unterschiedlich beurteilt, so war dies einmal ein Punkt für intraindividuelle Labilität, zum anderen aber auch für interindividuelle, da beide Urteile hier zusammen auftreten. Setzt man nun den Fall, daß es intraindividuelle Unterschiede gäbe, aber keine interindividuellen, d.h. daß zwar die einzelnen Vpn den 1. und 2. Durchgang unterschiedlich beurteilten, sie sich aber innerhalb eines Durchganges nicht voneinander unterschieden, so sollte ein Wert für die intraindividuelle Streuung herauskommen, aber die interindividuelle gleich Null sein. Faßt man aber beide Durchgänge zusammen, so erhält man für beide Labilitäten den gleichen Wert. Durch diese Vorgehensweise sind die beiden Labilitätswerte nicht voneinander unabhängig. Der intraindividuelle Wert wird zwar unabhängig ermittelt, der interindividuelle

hat aber in einem nicht abschätzbaren Ausmaß den ersteren in sich vereinigt und kann im Grenzfall, wenn keine Streuung vorliegt, nur genauso groß werden wie der intraindividuelle, ansonsten muß er größer sein.

Die BRÄUERschen Ergebnisse können in diesem Zusammenhang nur dahingehend interpretiert werden, daß es keinmal vorgekommen ist, daß keine interindividuelle Streuung vorlag.

3. Eine Modifizierung der BRÄUERschen Methode

Die Abhängigkeit des interindividuellen Labilitätswertes kann man ausschalten, wenn man die beiden Durchgänge nicht zusammenfaßt. Man berechnet den Labilitätswert des 1. und 2. Durchganges getrennt und bildet dann das arithmetische Mittel dieser Werte. Auf diese Weise erhält man ein interindividuelles Labilitätsmaß, daß nicht von dem intraindividuellen abhängig ist.

a. Die Versuche von BRÄUER

Um das unabhängige Maß für die interindividuelle Labilität zu erhalten, müssen die BRÄUERschen Daten noch einmal verrechnet werden. Die intraindividuellen Werte kann man übernehmen, da sie ja unabhängig von den interindividuellen ermittelt wurden. Bei den Versuchen mit drei Kategorien waren die intraindividuellen Werte stets gleich Null, auch hier muß man keine neuen Berechnungen anstellen, da diese Werte nicht die interindividuellen beeinflussen konnten. Berechnet man nun für die übrigen Versuche die interindividuellen Labilitätswerte, indem man das Mittel aus beiden Durchgängen heranzieht, so ergibt sich als Ergebnis die folgende Tabelle:

Tabelle 29

Labilitätswerte der BRÄUERschen Versuche nach der modifizierten Methode (ohne die Versuche mit 3 Kategorien)

Versuchs-Nr.	1.Durchgang	2.Durchgang	arithmetisches Mittel	BRÄUERschen Werte
7	5,9%	3,2%	4,6%	5,8%
9	4,3%	3,4%	3,9%	5,9%
1o	9,8%	9,8%	9,7%	1o,5%
11	9,4%	11,7%	1o,6%	12,4%
12	5,6%	5,1%	5,4%	5,8%
13	5,o%	6,1%	5,6%	5,8%
14	7,7%	6,o%	6,9%	7,6%
15	6,4%	4,9%	5,7%	5,9%
16	2,9%	2,7%	2,8%	3,1%
17	8,3%	9,6%	9,o%	1o,4%
19	5,2%	7,9%	6,6%	7,2%

Die nach der modifizierten Methode ermittelten interindividuellen Labilitätswerte liegen durchgängig etwas niedriger als die ursprünglichen. Setzt man nun diese Werte in Tabelle 28, Seite 98 ein, so ergibt sich folgendes Bild:

Tabelle 3o

Labilitätswerte der BRÄUERschen Versuche mit den modifizierten interindividuellen Werten

Versuchs-Nr.	interindivid. Labilität	intraindivid. Labilität	Anzahl der Kategorien	Alter	
6	o,2%	o,o%	3	14	
7	4,6%	2,8%	5	14	
8	o,6%	o,o%	3	11	
9	3,9%	5,5%	5	11	x
1o	9,7%	4,4%	5	9	
11	1o,6%	6,3%	5	1o	
12	5,4%	2,5%	5	11	
13	5,6%	1,8%	5	12	
14	6,9%	4,2%	5	13	
15	5,7%	2,9%	5	14	
16	2,8%	o,o%	5	Erwachsen	
17	9,o%	8,o%	5	11	
18	2,6%	o,o%	3	11	
19	6,6%	4,5%	5	14	
2o	o,7%	o,o%	3	14	

Durch die Korrektur der interindividuellen Werte ist auch der Fall eingetreten, daß einmal der interindividuelle Wert kleiner als der zugehörige intraindividuelle ausfällt (in der Tabelle durch ein Kreuz gekennzeichnet). Mit den Werten dieser Tabelle können wir nun den Wilcoxon-Test für Paardifferenzen durchführen. Dafür müssen die jeweiligen Differenzen zwischen inter- und intraindividuellem Wert berechnet, die Differenzen dem Vorzeichen entsprechend geordnet und die Rangplätze der Differenzen ermittelt werden. In unserem Fall haben wir nur eine negative Differenz beim 9. Versuch.

Die Berechnung mit Hilfe des Wilcoxon-Tests ergibt folgendes Ergebnis:

$$T_{pop} = \frac{N(N+1)}{4}, \qquad \sigma_T = \sqrt{\frac{(2N+1)}{6} \; T_{pop}}$$

$$z = (T_1 - T_{pop}) \; / \sigma_T$$

T_1 : Rangplatzsumme der negativen Differenzen (5)

N : Anzahl der Paare (15)

T_{pop} = 6o $\quad \sigma_T$ = 12,6 $\quad$ z = -4,4

Der zugehörige Wahrscheinlichkeitswert liegt zwischen 99,994 und 99,99994%.

Die Wahrscheinlichkeit, daß der Unterschied nur rein zufällig ist, beträgt weniger als o,o1%. Es zeigt sich also ein überaus signifikanter Unterschied zwischen den inter- und intraindividuellen Labilitätswerten dahingehend, daß die intraindividuellen überaus signifikant kleiner sind als die interindividuellen.

b. Die Versuche von HRUSCHKA

Nach der modifizierten Methode lassen sich nun auch die Labilitätswerte bei den von HRUSCHKA durchgeführten Versuchen ermitteln. Es ergibt sich hier jedoch eine neue Schwierigkeit.

Bei den Versuchen von BRÄUER gab es keine Systemenden. Die Versuche von HRUSCHKA weisen jedoch alle Systemenden auf. Daraus ergibt sich das Problem der Handhabung dieser Systemenden.

BRÄUER zählte die Punkte für Labilität aus und bezog sie auf die Zahl der insgesamt abgegebenen Urteile, die sich aus der Anzahl der dargebotenen Gegebenheiten mal der Anzahl der Vpn ergab. Die Anzahl der abgegebenen Urteile, da es keine Systemenden gab, also keine Gegebenheiten, die nicht mit einem kategorialen Urteil bedacht wurden. Bei

den Versuchen von HRUSCHKA ist diese Entsprechung nicht gegeben. Es gibt sehr wohl Gegebenheiten, die nicht mit einem kategorialen Urteil bedacht wurden, weil die Vpn diese als nicht mehr zum System gehörend angesehen haben.

Wie soll man diese Gegebenheiten handhaben? Es ergeben sich drei verschiedene Möglichkeiten, die alle zu unterschiedlichen Ergebnissen führen. Die Unterschiede sind zwar nur rein numerischer Natur, d.h. das Verhältnis zwischen inter- und intraindividueller Labilität wird davon nicht beeinträchtigt, jedoch die Vergleichbarkeit von Versuch zu Versuch, wenn man einmal Versuche mit und einmal Versuche ohne Systemenden vergleichen will.

Die erste zu treffende Entscheidung betrifft die Urteile, die nicht mehr zum System gehören. Sollen diese Urteile wie kategoriale Urteile aufgefaßt werden und mit in die Auszählung der Labilitätspunkte hereingenommen werden? Behandelt man die Systemenden wie eine Kategorie, so erhöht sich die Gesamtkategorienzahl des Versuches um 1 bei einem Systemende, um 2 bei beiden Systemenden.

Wählt man diese Vorgehensweise, so muß die Bezugszahl auch noch besonders beachtet werden. Man kann einmal den ganzen dargebotenen Bereich der Gegebenheiten mal der Anzahl der Vpn als Bezugszahl wählen. Hier wirkt sich dann aber die eigentliche Versuchskonzeption in der Größenordnung der Zahlen aus. Ist der Versuch so angelegt, daß sehr viele nicht zum System gehörende Gegebenheiten dargeboten werden, so ist die Bezugszahl recht groß und die resultierende Prozentzahl relativ klein. Wurde der Versuch aber so konzipiert, daß die Zahl der dargebotenen Gegebenheiten so gewählt wurde, daß zwar für jede Vpn Systemenden vorhanden sind, aber die Zahl der nicht mehr zum System gehörenden Gegebenheiten möglichst klein gehalten wurde, so ist die Bezugszahl relativ zum anderen Versuch kleiner und die resultierende Prozentzahl größer. Ein einfaches Zahlenbeispiel soll dies verdeutlichen:

Angenommen wir haben zwei völlig identische Versuche, die sich nur durch die Anzahl der nicht mehr zum System gehörenden Gegebenheiten unterscheiden. Bei Versuch 1 wurde die Gesamtzahl sehr großzügig festgelegt auf 8o Gegebenheiten, wobei viele Gegebenheiten abgelehnt wurden. Bei Versuch 2 wurde die Gesamtzahl der Gegebenheiten so gewählt, daß zwar echte Systemenden vorhanden sind, aber nicht viele Gegebenheiten nicht mehr akzeptiert wurden, die Gesamtzahl beträgt hier 4o.

Da sich beide Versuche ansonsten nicht unterscheiden sollen, wurde bei beiden insgesamt 15 Labilitätspunkte ausgezählt, die Anzahl der Vpn war 1o.

Die Berechnung der prozentualen Labilität ergibt folgende Prozentwerte:

Versuch 1: (15 x 1oo) / (1o x 8o) = o,94%
Versuch 2: (15 x 1oo) / (1o x 4o) = 1,87%

Bei der Größenordnung der Labilitätswerte kann dieser Unterschied schon eine beträchtliche Rolle spielen. Man kann sich nun behelfen, indem man nicht den vorgegebenen Bereich, also die Anzahl der insgesamt dargebotenen Gegebenheiten als Bezugszahl wählt, sondern selbst Grenzen zieht, die für alle Versuche gleich sind. Man kann von dem vorgegebenen Bereich alle die Gegebenheiten abziehen, die von allen Vpn als nicht zum System gehörend angesehen wurden, wodurch sich die Versuchskonzeption selbst nicht mehr auswirken kann. Man läßt den Bereich dort beginnen, wo das erste kategoriale Urteil erfolgt und ihn dort aufhören, wo das letzte kategoriale Urteil abgegeben wurde. Die Anzahl der Gegebenheiten, die in diesem Bereich liegen, mal der Anzahl der Vpn ist dann die Bezugszahl.

Geht man nun davon aus, daß die Systemenden nicht mehr zu den kategorialen Bereichen gehören, also nicht mit einbezogen werden sollen in die Auszählung der Labilitätspunkte, so dürfen diese Gegebenheiten

auch nicht mehr in der Bezugszahl auftauchen. Die Bezugszahl darf sich dann nur aus den wirklich kategorial abgegebenen Urteilen mal der Anzahl der Vpn zusammensetzen; man muß also für jede Vp auszählen, wie viele Urteile sie insgesamt kategorial abgegeben hat. Die Summe dieser Zahlen wird dann als Bezugszahl gewählt.

Die Versuche von HRUSCHKA wurden nach allen drei hier beschriebenen Möglichkeiten verrechnet und die folgende Tabelle zeigt, welche Ergebnisse dabei herausgekommen sind und wie sich die Ergebnisse entsprechend der Methode unterscheiden.

Tabelle 31

Die Labilitätswerte bei HRUSCHKA in Abhängigkeit zur Bezugsgröße

lfd. Nr. d.V.	Methode	Bezugsgrößen bei den Versuchsdurchgängen 1.D.	2.D.	3.D.	4.D.	5.D.	Labilität inter-indiv.	intra-indiv.	Anzahl der Kategor.
1	(ohne Systemenden)								
2	I	768	764	769	762	764	o,3%	o,1%	4
	II	81o	81o	81o	81o	81o	o,4%	o,2%	5
	III	81o	81o	81o	81o	81o	o,6%	o,4%	5
3	I	512	5oo				o,1%	o,o%	3
	II	144o	144o				o,5%	o,o%	3
	III	213o	213o				o,3%	o,o%	3
4	I	31o	3o6				5,4%	o,o%	3
	II	96o	96o				6,o%	o,o%	3
	III	111o	111o				5,o%	o,o%	3
5	I	4o8	397				o,1%	o,o%	3
	II	1o5o	1o5o				o,7%	o,o%	3
	III	123o	123o				o,6%	o,o%	3

Legende:

I: die Systemenden wurden weder für die Streuung mitbeachtet noch für die Bezugszahl verwendet, die Bezugszahl setzt sich aus den kategorial abgegebenen Urteilen zusammen.

II: die Systemenden wurden als Kategorie angesehen und für die Bezugszahl wurde der Bereich gewählt, in dem zu jeder Gegebenheit mindestens ein kategoriales Urteil abgegeben wurde.

III: die Systemenden wurden als Kategorie angesehen und für die Bezugszahl wurde der gesamte dargebotene Bereich der Gegebenheiten gewählt.

Die Tabelle zeigt, daß es teilweise zu recht unterschiedlichen Ergebnissen bei den einzelnen Verrechnungsarten kommt, bei Versuch 5 unterscheiden sich die Werte z.B. recht deutlich: läßt man die Systemenden völlig unbeachtet, so ist der interindividuelle Labilitätswert o,1%, betrachtet man die Systemenden wie eine Kategorie, so ist der Labilitätswert o,7% oder o,6%. Die beiden Möglichkeiten bei Miteinbeziehung der Systemenden unterschieden sich untereinander weniger.

Will man die Werte von HRUSCHKA mit denen von BRÄUER vergleichen, muß man sich entscheiden, welche Verrechnungsart als angemessen angesehen werden soll. Die dritte Art, bei der die Systemenden als Kategorien betrachtet wurden und die Gesamtzahl der dargebotenen Gegebenheiten als Bezugszahl diente, kann von vorneherein ausgeschlossen werden, da sie zu sehr von der Willkür der jeweiligen Versuchskonzeption abhängt. Die Entscheidung zwischen der ersten und zweiten Verrechnungsart ist nicht so einfach zu fällen. Dazu bedarf es noch einer eingehenden Betrachtung der erzielten Ergebnisse.

c. Unterschiede entsprechend der Kategorienzahl

Betrachtet man nur die intraindividuellen Labilitätswerte, so fiel schon bei BRÄUER auf, daß die Labilität bei Versuchen mit drei Kategorien stets gleich Null war. Diese Feststellung trifft auch für die Ergebnisse der Versuche von HRUSCHKA zu. Auch hier sind die Labilitätswerte bei Versuchen mit drei Kategorien stets gleich Null. Sie sind aber auch dann gleich Null, wenn die Systemenden bei der Verrechnung mit einbezogen wurden, wodurch sich die Anzahl der Kategorien ja eigentlich auf fünf erhöht. Bei Versuchen mit fünf Kategorien gab es aber stets eine intraindividuelle Labilität. Aufgrund dieser Tatsache kann man vermuten, daß die Systemenden für die Vpn inhaltlich eine andere Bedeutung haben als die Kategorien, daß die kategorialen Urteile von anderen Bedingungsgefügen beeinflußt werden als die Urtele 'nicht mehr zum System gehörend'.

Die unterschiedlichen Beeinflussungsgefüge bei den beiden Urteilsarten zu untersuchen, wird bestimmt einige interessante Aspekte für die Bezugssystemforschung erbringen. Da diese Untersuchungen jedoch bislang noch nicht vorgenommen wurden, bleibt in diesem Zusammenhang nur der Schluß, daß man für den Vergleich der Streuung der Urteile am besten die Systemenden nicht mit einbezieht, da hier wahrscheinlich Variablen zur Wirkung kommen, deren Bedeutung und Erstreckung noch nicht bekannt ist.

Für die Zusammenstellung aller Versuchsergebnisse zum Problem der Streuungsunterschiede können wir nur die Ergebnisse von HRUSCHKA heranziehen, die ohne Berücksichtigung der Systemenden gewonnen wurden. Die folgende Tabelle gibt das Gesamt der Ergebnisse von HRUSCHKA und BRÄUER wieder, wobei nach der Anzahl der im Versuch verwendeten Kategorien unterschieden wird.

Tabelle 32

Labilitätswerte geordnet nach der Kategorienzahl

3 Kategorien		4 Kategorien		5 Kategorien	
inter	intra	inter	intra	inter	intra
o,2	o,o	o,3	o,1	4,6	2,8
o,6	o,o			3,9	5,5
2,6	o,o			9,7	4,4
o,7	o,o			1o,6	6,3
o,1	o,o			5,4	2,5
5,4	o,o			5,6	1,8
o,1	o,o			6,9	4,2
				5,7	2,9
				2,8	o,2
				9,o	8,o
				6,6	4,5
				o,4	o,2

Vergleicht man die Labilitätswerte unter dem Gesichtspunkt der Kategorienanzahl des einzelnen Versuches, so steigen die Werte mit zunehmender Kategorienzahl. Die intraindividuellen Werte sind bei 3 Kategorien stets gleich Null, bei dem einen Versuch mit 4 Kategorien etwas größer als Null (o,1) und bei den Versuchen mit 5 Kategorien durchgängig beträchtlich größer als Null. Die interindividuellen Werte sind immer größer als Null, aber auch hier läßt sich eine steigende Tendenz mit zunehmender Kategorienzahl erkennen. Bei den Versuchen mit 3 Kategorien liegen die Werte ziemlich nahe bei Null, nur bei einem Versuch liegt der Labilitätswert mit 5,4 sehr hoch. Bei den Versuchen mit 5 Kategorien liegen die Werte häufig sehr weit von Null entfernt und nur in Ausnahmefällen dicht bei Null (o,4).

Es wäre nun interessant zu sehen, wie sich die Verhältnisse ändern, wenn man Versuche mit 6 oder 7 Kategorien zum Vergleich hätte. Der Unterschied zwischen den interindividuellen Werten bei Versuchen mit 3 und mit 5 Kategorien ist überaus signifikant (Berechnet mit Hilfe des t-Tests: $t = 4,2$ $df = 17$, ein t von 2,898 ist bereits auf dem 1%-Niveau signifikant).

Es stellt sich nun die Frage, warum sich die einzelnen Labilitätswerte bei unterschiedlicher Kategorienzahl unterscheiden oder warum die Labilitätswerte steigen, wenn die Anzahl der Kategorien größer wird?

Gehen wir von dem von BRÄUER festgelegten Maß für einen Labilitätspunkt aus. Labilität liegt danach dann vor, wenn eine Gegebenheit mit mehreren Kategorien beurteilt wird, die mehr als eine Kategorie auseinander liegen. Bei Versuchen mit drei Kategorien kann dieser Fall nur dann erreicht werden, wenn eine Gegebenheit mit beiden Extremkategorien bedacht wird: bei den Kategorien groß, mittel und klein also mit der Kategorie groß und klein. Bei Versuchen mit fünf Kategorien kann dieser Fall viel öfter eintreten: bei den Kategorien sehr groß, groß, mittel, klein und sehr klein dann, wenn eine Gegebenheit mit den Kategorien sehr groß und mittel, groß und klein oder mittel und sehr klein beurteilt wurde.

Unterschiede über mehr als zwei Kategorien kann es bei Versuchen mit drei Kategorien gar nicht geben, bei Versuchen mit fünf Kategorien sind diese jedoch möglich: sehr groß-klein, groß-sehr klein, sehr groß-sehr klein. Insgesamt gibt es also bei Versuchen mit drei Kategorien nur eine einzige Möglichkeit, einen Labilitätspunkt zu erlangen, bei Versuchen mit fünf Kategorien bereits 6 Möglichkeiten.

Abgesehen von den Möglichkeiten haben die Unterschiede bei Versuchen mit 3 oder 5 Kategorien eine andere inhaltliche Bedeutung. Unterscheidet

sich eine Gegebenheit bei einem Versuch mit drei Kategorien bei beiden Urteilen um mehr als eine Kategorie, so wurde bei den beiden Urteilen jeweils der Extrempunkt der kategorialen Skala gewählt: groß und klein. Bei Versuchen mit 5 Kategorien muß sich der Unterschied jedoch nicht auf die Extremklassen beziehen. Wird hier eine Gegebenheit einmal mit sehr groß und dann mit mittel beurteilt, so bedeutet dies etwas anderes, als wenn eine Gegebenheit bei insgesamt drei Klassen mit groß und klein beurteilt wurde. Wird bei den Versuchen mit 5 Kategorien eine Gegebenheit mit den Extremklassen bedacht, so unterscheiden sich die Urteile bereits um 4 Kategorien. Die Markierung "Unterschied über mehr als eine Kategorie" bedeutet für die Versuche mit verschiedener Kategorienzahl etwas anderes. Es ist deshalb nicht verwunderlich, daß die Labilitätswerte mit zunehmender Kategorienzahl ebenfalls größer werden.

Die von BRÄUER gewählte Einteilung zeigt sich bereits für die intraindividuelle Labilität bei Versuchen mit drei Kategorien als nicht ausreichend, da stets keine Labilitäten dadurch auftreten. Man müßte dieses Maß entsprechend der Kategorienzahl differenzieren. Bei weniger Kategorien sollte man ein groberes Maß nehmen als bei vielen Kategorien. Mit dem bisherigen Maß ist auf jeden Fall ein Vergleich von Versuchen mit unterschiedlicher Kategorienzahl nicht möglich. Dies bedeutet, daß man den Wilcoxon-Test nicht über alle Versuche anwenden darf, sondern nur getrennt für Versuche mit gleicher Kategorienzahl. Berechnet man den Wilcoxon-Test nach diesen Überlegungen noch einmal, so erhält man einen z-Wert von -3, der bei einem N von 12 eine Wahrscheinlichkeit von 99,73% hat, d.h. die Wahrscheinlichkeit, daß die beiden Wertereihen intraindividuell und interindividuell sich nur zufällig unterscheiden ist gleich o,27%. Das Ergebnis bestätigt die Hypothese, es beruht jedoch nur auf den Werten von Versuchen mit 5 Kategorien, da das angewendete Maß für Versuche mit drei Kategorien zu grob ist.

Zur besseren Ausschöpfung der Ergebnisse wäre eine Methode wünschenswert, die nicht von der Anzahl der Kategorien abhängt und die auch bei den Versuchen mit weniger Kategorien differenzierte Ergebnisse liefern kann.

4. Die Methode von Mittelwert und Streuung

Die bisher angewendeten Methoden für die Berechnung des Streuungsvergleiches erwiesen sich beide als mehr oder minder unzulänglich. Geht man von der Aussage der zu untersuchenden Hypothese aus, so liegt eigentlich die Berechnung von Mittelwerten und den zugehörigen Streuungen nahe. Das Problem ist hierbei nur, daß wir für die Erstellung eines Mittelwertes numerische Daten benötigen. Die Versuchsergebnisse selbst liegen uns jedoch nicht in Form von numerischen Daten vor, sondern wir haben lediglich die Urteile zu einzelnen Gegebenheiten. Der Vorteil der BRÄUERschen Methode liegt darin, daß sie direkt auf die vorliegenden Ergebnisse angewendet werden kann, sie benötigt keine numerischen Ausgangsdaten.

Wollen wir Mittelwerte berechnen, so müssen wir die primären Versuchsergebnisse irgendwie in numerische Daten transformieren, wodurch auf jeden Fall ein Informationsverlust bewirkt wird. Zählen wir für die einzelnen Vpn die Urteile aus, die in einer bestimmten Kategorie abgegeben wurden, so ist damit der Bezug der Urteile zu den zugrundeliegenden Gegebenheiten verloren gegangen. Wir können an den Zahlen selbst nicht einmal mehr erkennen, wo im Bereich der dargebotenen Gegebenheiten diese Kategorie gelegen hat. Berechnen wir also die Mittelwerte über die Größe der einzelnen Kategorien und die zugehörigen Streuungen, so müßte die Hypothese entsprechend abgeändert werden: Die Anzahl der Urteile in einer bestimmten Kategorie unterscheidet sich intraindividuell überaus signifikant weniger als interindividuell. Anders ausgedrückt, wählt eine Vp eine Kategorie bei mehreren Versuchsdurchgängen eher gleich häufig als verschiedene Vpn bei einem Versuchsdurchgang. Die Ausgangshypothese spricht aber nicht von der Häufigkeit, in der eine Kategorie gewählt wurde, sondern ganz explizit von den Urteilen selbst.

Wollen wir die Information, welche Gegebenheit mit einer bestimmten

Kategorie bedacht wurde, mit in den Zahlenwert hineinbringen, so dürfen wir nicht von der Anzahl der Urteile in einer Kategorie alleine ausgehen. Betrachten wir vielmehr den Punkt, wo die Urteile von einer Kategorie in die nächste umschlagen, so haben wir damit in etwa eine Information darüber, in welchem Bereich der vorgegebenen Gegebenheiten sich die Urteile bewegen.

Wir können zwar immer noch nicht mit diesem Wert eine Zuordnung Gegebenheit - Urteil erreichen, wir können aber ungefähr abstecken, welche Gegebenheiten mit welchen Urteilen bedacht werden. Der Punkt, an dem die Urteile von einer Kategorie zur nächsten überwechseln wird in der Bezugssystemforschung mit Zäsur bezeichnet. Die Zäsur ist quasi die Grenze zwischen zwei aneinanderstoßenden Kategorien. Untersuchen wir Unterschiede bei den Zäsuren, so berücksichtigen wir nicht allein die Anzahl der Urteile, die zu einer Kategorie abgeben wurden, sondern haben ebenfalls eine Information darüber, in welchem Bereich der Skala der dargebotenen Gegebenheiten diese Kategorie zu finden ist. Beurteilt eine Vp bei zwei Versuchsdurchgängen die Gegebenheiten mit unterschiedlichen Kategorien, gibt aber insgesamt gesehen in jeder Kategorie in etwa gleich viele Urteile ab, so haben sich die Zäsuren verschoben. Berechnet man nun die Mittelwerte für diese Zäsuren und die zugehörigen Streuwerte, so erhält man durchaus eine Information über die Unterschiede bezüglich einzelner Beurteilungen. Diese Streuwerte können somit als Ausdruck der in der Hypothese angesprochenen Streuung angesehen werden.

Wird ein Versuch zweimal durchgeführt, so erhält man für jede Vp jeweils zwei Zäsurwerte am Übergang von einer Kategorie zur anderen. Man rechnet für diese beiden Werte die Streuung aus, berechnet diese Streuung für alle Vpn und mittelt dann die gefundenen Streuwerte; so erhält man ein Maß für die intraindividuelle Streuung. Berechnet man die Streuung der Zäsuren bei einem Versuchsdurchgang über alle Vpn und mittelt die gefundenen Werte der einzelnen Durchgänge, so erhält man ein Maß für die interindividuelle Streuung.

Die Zäsurwerte selbst sind dabei nicht abhängig von der Anzahl der verwendeten Kategorien pro Versuch, man erhält jedoch entsprechend der Kategorienanzahl unterschiedlich viele Zäsuren: Bei drei Kategorien erhält man zwei Zäsuren, zwischen 1. und 2. und zwischen 2. und 3. Kategorie, bei Versuchen mit 4 Kategorien 3 Zäsuren und bei Versuchen mit 5 Kategorien 4 Zäsuren.

Die folgende Tabelle faßt die Ergebnisse dieser Berechnungen zusammen. Die interindividuellen Streuwerte sind dabei einmal gemittelt aufgeführt und danach getrennt für jeden Versuchsdurchgang, um eventuelle Schwankungen von Versuchsdurchgang zu Versuchsdurchgang darzulegen. (Tabelle 33 folgende Seite 116)

Die ermittelten Streuwerte erstrecken sich im Sinne der Hypothese, die intraindividuellen Werte liegen durchgängig niedriger als die zugehörigen interindividuellen Werte. Im Gegensatz zu den vorher beschriebenen Methoden kann man außerdem noch die Bereiche der interindividuellen und intraindividuellen Streuwerte klar voneinander abgrenzen. Der Bereich der intraindividuellen Werte erstreckt sich von o,5 bis 2,5, der der interindividuellen Werte von 3,o bis 9,3. Es gibt also keine Überschneidungen, das Gesamt der intraindividuellen Werte liegt niedriger als das Gesamt der interindividuellen Werte und nicht nur die zugehörigen Paare. Der Gesamtstreubereich der beiden Gruppen unterscheidet sich auch noch: die intraindividuellen Werte streuen insgesamt weniger, nur von o,5 bis 2,5, die interindividuellen jedoch wesentlich mehr, von 3,o bis 9,3.

Die Ergebnisse sind so eindeutig, daß sich eine Überprüfung mit dem Wilcoxon-Test erübrigt, er kann nur 1oo igen Unterschied ergeben.

Beleuchtet man die Streuungsunterschiede innerhalb der Ergebnisse, so unterscheiden sich die interindividuellen Streuungen in den einzelnen Durchgängen nur geringfügig:

mittlere Streuung 1. Durchgang: 5,2
mittlere Streuung 2. Durchgang: 5,8

Die Tabelle 34 stellt die Ergebnisse noch einmal dar, jedoch nach den Zäsuren geordnet und ohne Zwischenergebnisse (Seite 117).

Tabelle 33

Streuungen der Zäsuren

Lfd. Nr.	Zäsur	intra	inter A.M.	interindividuell 1.Dg.	2.Dg.	3.Dg.	4.Dg.	5.Dg.
1	1	o.9	3.2	3.4	3.o			
	2	1.5	5.5	4.6	6.3			
	3	1.2	5.4	5.3	5.5			
2	1	1.6	3.o	3.6	2.1	3.5	3.5	2.5
	2	2.o	4.o	4.2	4.1	4.2	4.1	3.2
	3	2.5	3.6	3.3	3.o	3.8	3.8	4.2
	4	1.6	4.7	5.o	3.9	5.2	4.7	4.4
3	1	o.8	3.6	3.9	3.9			
	2	o.8	4.6	4.3	4.8			
4	1	1.1	5.5	5.1	5.9			
	2	1.4	4.5	4.6	4.3			
5	1	1.1	4.8	4.4	5.1			
	2	1.3	6.1	4.5	7.6			
6	1	1.2	6.3	5.8	6.7			
	2	1.o	8.5	8.1	8.8			
7	1	1.6	6.8	6.4	7.2			
	2	1.2	8.7	8.8	8.6			
	3	1.4	6.4	6.8	5.9			
	4	o.6	3.2	3.6	2.8			
8	1	1.1	7.4	7.4	7.4			
	2	o.9	4.5	4.7	4.2			
9	1	1.5	3.6	3.6	3.5			
	2	1.8	5.8	4.6	7.o			
	3	1.8	5.9	6.1	5.7			
	4	o.9	3.o	2.8	3.2			
1o	1	o.9	5.5	5.4	5.5			
	2	1.5	8.5	7.9	9.1			
	3	1.2	8.o	7.8	8.2			
	4	1.6	9.3	8.9	9.7			
11	1	1.2	6.6	6.9	6.3			
	2	1.7	8.9	9.1	8.7			
	3	1.6	6.7	5.7	7.6			
	4	1.1	7.2	6.5	7.8			
12	1	1.1	3.2	3.o	3.3			
	2	1.3	5.9	4.6	7.2			
	3	1.o	5.1	4.9	5.2			
	4	1.2	5.2	4.2	6.1			
13	1	o.7	4.8	4.4	5.1			
	2	o.7	5.7	5.6	5.7			
	3	1.o	4.4	4.4	4.3			
	4	1.o	6.7	6.o	7.3			
14	1	o.5	4.3	4.4	4.2			
	2	1.1	6.7	6.3	7.1			
	3	1.2	5.3	4.7	5.9			
	4	1.2	7.2	8.1	6.3			

lfd. Nr.	Zäsur	intra	inter A.M.	inter. 1.Dg.	2.Dg.
15	1	o.9	4.2	3.5	4.9
	2	1.5	8.8	8.2	9.4
	3	1.1	5.2	4.o	6.2
	4	1.o	5.3	6.1	4.4
16	1	o.7	5.o	4.5	5.5
	2	o.9	5.3	5.o	5.5
	3	1.o	4.2	3.6	4.8
	4	1.1	5.1	5.2	4.9
17	1	1.3	6.2	6.9	5.5
	2	1.2	3.9	3.1	4.7
	3	1.7	6.o	5.3	6.6
	4	1.4	8.o	8.3	7.6
18	1	2.5	5.9	5.2	6.5
	2	1.o	4.3	5.2	3.3
19	1	1.9	6.o	4.3	7.6
	2	1.2	4.8	3.6	7.o
	3	o.9	4.2	3.9	4.5
	4	o.7	4.o	3.9	4.o
2o	1	1.3	3.7	2.6	4.7
	2	o.8	4.1	4.1	4.1

Tabelle 34

Streuung der Zäsuren
(ohne die Ergebnisse der einzelnen Durchgänge)

lfd. Nr.	1.Zäsur intra	1.Zäsur inter	2.Zäsur intra	2.Zäsur inter	3.Zäsur intra	3.Zäsur inter	4.Zäsur intra	4.Zäsur inter
1	o.9	3.2	1.5	5.5	1.2	5.4		
2	1.6	3.o	2.o	4.o	2.5	3.6	1.6	4.7
3	o.8	3.6	o.8	4.6				
4	1.1	5.5	1.4	4.5				
5	1.1	4.8	1.3	6.1				
6	1.2	6.3.	1.o	8.5				
7	1.6	6.8	1.2	8.7	1.4	6.4	o.6	3.2
8	1.1	7.4	o.9	4.5				
9	1.5	3.6	1.8	5.8	1.8	5.9	o.9	3.o
1o	o.9	5.5	1.5	8.5	1.2	8.o	1.6	9.3
11	1.2	6.6	1.7	8.9	1.6	6.7	1.1	7.2
12	1.1	3.2	1.3	5.9	1.o	5.1	1.2	5.2
13	o.7	4.8	o.7	5.7	1.o	4.4	1.o	6.7
14	o.5	4.3	1.1	6.7	1.2	5.3	1.2	7.2
15	o.9	4.2	1.5	8.8	1.1	5.2	1.o	5.3
16	o.7	5.o	o.9	5.3	1.o	4.2	1.1	5.1
17	1.3	6.2	1.2	3.9	1.7	6.o	1.4	8.o
18	2.5	5.9	1.o	4.3				
19	1.9	6.o	1.2	4.8	o.9	4.2	o.7	4.o
2o	1.3	3.7	o.8	4.1				

Mittlere Streuwerte der Zäsuren

	interindividuell	intraindividuell
1. Zäsur	4.9	1.2
2. Zäsur	5.9	1.2
3. Zäsur	5.4	1.4
4. Zäsur	5.7	1.1

Berechnet man für die einzelnen Zäsuren getrennt die mittleren Streuwerte, so ergeben sich auch hier nur geringfügige Unterschiede. Das Gesamt der Ergebnisse erstreckt sich in einem relativ homogenen Feld, Einflüsse durch die Anzahl der verwendeten Kategorien bei den einzelnen Versuchen lassen sich nicht erkennen. Bedenkt man, daß die Ergebnisse von recht unterschiedlichen Versuchen herstammen, so ist es erstaunlich, wie diese Verrechnungsmethode zu insgesamt sehr einheitlichen und somit gut vergleichbaren Ergebnissen führt.

II. Methodenvergleich, Zusammenfassung der Ergebnisse und Ausblick auf spätere Versuche

1. Zusammenfassung

In dem vorangegangenen Abschnitt wurde die Hypothese über die Streungsunterschiede behandelt. Zu überprüfen war die Aussage, daß die intraindividuellen Streuungen überaus signifikant kleiner seien als die interindividuellen.

Für die Behandlung der Hypothese konnte kein fertiges statistisches Verfahren herangezogen werden. Die Art der Fragestellung und der vorhandenen Versuchsergebnisse erforderte eine spezielle Methode, die noch gefunden werden mußte. In einer früheren Arbeit wurde in einem anderen Zusammenhang bereits so etwas wie ein Streuvergleich vorgenommen. Zum Ausgangspunkt der Bearbeitung wurde deshalb die dort praktizierte Methode gewählt:

BRÄUER ging in seiner Arbeit direkt von den Versuchsergebnissen aus, und zwar von den einzelnen Urteilen zu den jeweiligen Gegebenheiten. Er betrachtete einen Beurteilungsunterschied dann als eine Streuung oder Labilität, wenn sich die Urteile um mehr als eine Kategorie unterschieden.

Er zählte diese Unterschiede zusammen und setzte sie in Bezug zu der Gesamtzahl der abgegebenen Urteile. Für die intraindividuelle Streuung betrachtete er die Urteile einer Vp vom ersten und zweiten Durchgang und berechnete einen Labilitätspunkt, wenn sich diese beiden Urteile um mehr als eine Kategorie unterschieden. Diese Prüfung führte er für jede Vp durch und setzte die Summe der Labilitätspunkte in Bezug zu den abgegebenen Urteilen oder den dargebotenen Gegebenheiten, die gleich derjenigen der abgegebenen Urteile waren, da BRÄUER ohne Systemenden arbeitete.

Für die interindividuelle Streuung zählte er die Häufigkeit der Urteile in den einzelnen Kategorien bei jeder Gegebenheit (Quersumme gleich Anzahl der Vpn) zusammen, bestimmte die häufigste Kategorie als Maßstab und zählte, wie oft sich die Urteile von dieser häufigsten Kategorie um mehr als eine Kategorie unterschieden. Die Summe dieser Labilitätspunkte über alle Gegebenheiten bezog er wiederum auf die Anzahl der insgesamt abgegebenen Urteile und erhielt so ein interindividuelles Streumaß, das mit dem intraindividuellen verglichen werden konnte. BRÄUER unterschied bei dieser Prozedur jedoch nicht zwischen dem ersten und zweiten Versuchsdurchgang, die Häufigkeit der Urteile pro Gegebenheit war bei ihm 2 mal der Anzahl der Vpn, wodurch bereits intraindividuell sich auswirkende Labilitätspunkte in den interindividuellen Wert mit eingingen und beide Werte nicht voneinander unabhängig waren. Im Zusammenhang dieser Untersuchung war diese Tatsache nicht tragbar, da es ja gerade auf den Vergleich der beiden Werte ankam. Die Methode wurde deshalb modifiziert, es wurden die interindividuellen Werte für jeden Versuchsdurchgang getrennt berechnet und dann gemittelt, wodurch die Abhängigkeit der beiden Werte inter und intra aufgehoben wurde. Bei der Anwendung dieser Methode auf die Versuche von HRUSCHKA ergaben sich Schwierigkeiten, da diese Versuche mit Systemenden konzipiert waren und die Systemenden einmal unbeachtet gelassen, zum anderen aber auch als Kategorien aufgefaßt werden konnten. Bei dem Vergleich der Ergebnisse zeigte sich jedoch, daß die Systemenden nicht als Kategorien aufzufassen sind, da ihren Beurteilungen wahrscheinlich andere Bedingungsgefüge zugrundeliegen als den kategorialen Beurteilungen.

Die Ergebnisse, die aufgrund dieser Verrechnungsweise erzielt wurden, lassen sich folgendermaßen zusammenfassen:

1. Bei Versuchen mit drei Kategorien zeigten sich durchweg keine intraindividuellen Labilitäten in dem von BRÄUER definierten Sinne, interindividuelle Labilitäten waren vorhanden.

2. Bei den Versuchen mit fünf Kategorien gab es inter- und intraindividuelle Labilitäten, wobei die intraindividuellen signifikant kleiner waren.

3. Es zeigte sich eine Tendenz dahingehend, daß mit zunehmender Kategorienzahl die Anzahl der Labilitätspunkte zunahm.

Das dritte Ergebnis ist auf eine Abhängigkeit der Methode von der Anzahl der zu verwendenden Kategorien zurückzuführen. Der von BRÄUER definierte Unterschied, der zu einem Labilitätspunkt führt, ist einmal rein von der Kombinatorik her bei Versuchen mit weniger Kategorien weniger möglich als bei Versuchen mit mehr Kategorien. Außerdem bedeutet die definierte Urteilsabweichung bei weniger Kategorien inhaltlich eine viel größere Abweichung als bei mehr Kategorien. Für Versuche mit drei Kategorien zeigte sich das Verfahren als zu grob, intraindividuelle Unterschiede konnte es bereits nicht mehr anzeigen. Durch die Abhängigkeit von der Kategorienzahl ist ein Vergleich von Versuchen mit unterschiedlicher Kategorienzahl nicht möglich.

Auf Grund dieser Tatsache wurde noch eine weitere Methode für die Bearbeitung der Streuungshypothese angewendet. Da die Hypothese selbst von Streuung spricht, lag es nahe, auf irgendeine Art eine Streuung zu berechnen. Mittelwerte und zugehörige Streuungen können jedoch nur bei numerischen Daten berechnet werden. Die Umwandlung des vorhandenen Datenmaterials in numerische Zahlen bedeutete auf jeden Fall einen Informationsverlust, die Zuordnung Urteil - Gegebenheit konnte nicht mehr vollzogen werden. Um ein möglichst hohes Maß an Information in die numerischen Werte einzubeziehen, wurden die Zäsuren berechnet, die global eine Zuordnung der Urteile in einer Kategorie zu einem Bereich der dargebotenen Gegebenheiten erlauben. Die Zäsuren selbst sind numerische Werte, mit denen man Mittelwert und Streuung berechnen kann. Diese Vorgehensweise erbrachte folgende Ergebnisse:

1. Die intraindividuelle Streuung ist immer kleiner als die interindividuelle.

2. Der Bereich, in dem sich die intraindividuellen Streuwerte bewegen, liegt niedriger als der Bereich der interindividuellen.
3. Das Gesamt der Ergebnisse ist ausgesprochen homogen verteilt, Unterschiede aufgrund der Kategorienzahl lassen sich nicht feststellen, eine Vergleichbarkeit über alle Versuche hinweg ist gegeben.

2. Vergleich der beiden Methoden

Die Ergebnisse, die mit beiden Methoden, der BRÄUERschen und der von Mittelwert und Streuung, erzielt wurden, sind ihrer Tendenz nach gleich, beide zeigen den in der Hypothese beschriebenen Unterschied zwischen inter- und intraindividueller Streuung. Die Ergebnisse selbst aber unterscheiden sich teilweise recht deutlich. Mit Hilfe der BRÄUERschen Methode waren die intraindividuellen Werte zwar meist kleiner als die zugehörigen interindividuellen Werte, die beiden Bereiche überlappten sich aber recht häufig, wenn man das Gesamt der inter- und intraindividuellen Werte heranzieht. Bei der Methode von Mittelwert und Streuung konnte man die beiden Bereiche klar voneinander abgrenzen, sie überschnitten sich nie und der interindividuelle Bereich wies eine größere Streuung seiner Werte auf als der intraindividuelle.

Es stellt sich nun die Frage, ob diese Unterschiede nur im reinen Zahlenbereich liegen, daß die Skalen sich in unterschiedlicher Weise erstrecken, oder ob die Ergebnisse gar nicht miteinander vergleichbar sind und verschiedene Aspekte zur Grundlage haben.

Da die BRÄUERsche Methode von den vorliegenden Versuchsbedingungen abhängig ist, wird für den Methodenvergleich eine Versuchsreihe herausgegriffen, bei der nur das Alter variiert, die Gegebenheiten, deren Gesamtanzahl und die Anzahl der Kategorien jedoch konstant sind. Nach der BRÄUERschen Methode erhalten wir für jeden Versuch ein Wertepaar, nach der Methode von Mittelwert und Streuung so viele Wertepaare,

wie Zäsuren vorhanden sind. Um vergleichbare Werte zu erhalten, werden die einzelnen Zäsurwerte gemittelt. Die folgende Tabelle zeigt die Wertepaare dieser Versuchsreihe einmal nach der BRÄUERschen Methode und zum anderen nach der Streumethode:

Tabelle 35

Inter- und intraindividuelle Streuungen
Versuchsreihe:
BRÄUER, Versuche mit 54 Bleistiften, 5 Kategorien zur Beurteilung und Vpn im Alter von 9 bis 14 Jahren und Erwachsenen.

Alter	9	1o	11	12	13	14	Erw.
BRÄUERsche Meth.							
interindividuell	9,7	1o,6	5,4	5,6	6,9	5,7	2.8
intraindividuell	4,4	6,3	2,5	1,8	4,2	2,9	o,2
Meth. von Mittelwert und Streuung							
interindividuell	7,8	7,4	4,9	5,4	5,9	5,9	4,9
intraindividuell	1,3	1,4	1,2	o,9	1,o	1,1	o,9

Diese Versuchsreihe von BRÄUER sollte altersbedingte Unterschiede der Stabilität der Urteile anzeigen, einmal bei den einzelnen Vpn selbst (intra), zum anderen bei der Altersgruppe (inter). BRÄUER kommt zu dem Schluß, daß die Werte von 9 bis 12 Jahren eine abfallende Tendenz, also eine Tendenz zur Stabilität hin aufweisen, danach während der Pubertät oszilieren und beim Erwachsenen das größte Maß an Stabilität oder die geringste Labilität erreichen.

Vergleicht man die Werte nach beiden Methoden unter diesem Gesichts-

punkt, so ergibt sich kein grundlegender Unterschied, auch die unteren beiden Wertereihen zeigen die von BRÄUER beschriebene Tendenz.

Ein Unterschied ist eigentlich nur bei den intraindividuellen Werten bezüglich ihrer Größenordnung zu finden. Bei BRÄUER haben die intraindividuellen Werte einen recht großen Streubereich, sie reichen von 4,4 bis o,2. Die Werte nach der anderen Methode reichen nur von 1,4 bis o,9, einem sehr kleinem Streubereich, in dem ein Unterschied von o,1 bereits beträchtlich ist. Die Bereiche der interindividuellen Werte liegen in etwa gleich, wenn auch hier die BRÄUERschen Werte mehr streuen. Diese Unterschiede in den Zahlenbereichen und der Streuung werden Folgen der verschiedenen Berechnungsmodi sein, inhaltliche Unterschiede lassen sich nicht erkennen.

3. Ausblick auf spätere Untersuchungen

Die hier angestellten Analysen haben alle die in der zugrundeliegenden Hypothese gemachten Aussagen bestätigt. Es ist somit als bewiesen anzusehen, daß die intraindividuelle Streuung signifikant kleiner ist als die interindividuelle.

Es wäre in diesem Zusammenhang vielleicht recht interessant, die BRÄUERsche Methode dahingehend zu verändern, daß die Ergebnisse nicht mehr von der Anzahl der im Versuch verwendeten Kategorien beeinflußt werden. Man sollte versuchen, ein der Kategorienzahl angepaßtes Maß für Labilität zu finden, um damit die Vergleichbarkeit von Versuch zu Versuch wieder herzustellen. Die BRÄUERsche Methode völlig beiseite zu legen, da mit der Methode von Mittelwert und Streuung "bessere" Ergebnisse erzielt werden, ist nicht richtig, weil die zweite Methode durch die Umwandlung der Daten einen Informationsverlust bedingt. Nur die Methode von BRÄUER ist bislang in der Lage, die gesamte Information des Versuchs einzubeziehen und sollte deshalb erweitert werden. Hat

man z.B. eine andere Fragestellung als bei dieser Untersuchung, möchte man Unterschiede bei einzelnen Versuchsarten oder Gegebenheiten feststellen, kann die zweite Methode leicht bestehende Unterschiede verwischen, da sie allgemein zu relativ homogenen Ergebnissen führt.

Ein weiterer interessanter Gesichtspunkt dürfte die Art der Systemenden sein. Festgestellt wurde hier lediglich, daß sie nicht wie die Kategorien zu behandeln sind. Untersucht werden muß noch, welche Variablen bei den kategorialen Urteilen eine Rolle spielen und welche demgegenüber bei den Urteilen 'nicht mehr zum System gehörend' von Wichtigkeit sind.

Dritter Teil:

Zusammenfassung

Die vorliegende Arbeit hatte die 'mathematische Prüfbarkeit und Prüfung von Hypothesen über mnestisch stabilisierte Bezugssysteme' zum Thema. Es ging dabei weniger um eine experimentelle Überprüfung dieser Hypothesen, sondern um den Versuch, sie anhand der bislang in der Bezugssystemforschung unternommenen Untersuchungen zu überprüfen. Da die Hypothesen jedoch eine Folgerung dieser Versuche darstellen und nicht die Versuche in Hinblick auf diese Hypothesen unternommen worden sind, ergab sich sofort das Problem, wie diese Versuchsdaten mathematisch zu verarbeiten wären. Eine mathematisch-statistisch exakte Überprüfung war kaum möglich und im Mittelpunkt der Arbeit standen deshalb Überlegungen für spätere, experimentelle Untersuchungen dieser Hypothesen, Überlegungen über Abhängigkeitsverhältnisse und Beeinflussungsmöglichkeiten.

Die sechs hier zur Diskussion stehenden Hypothesen wurden von WITTE aufgestellt und umfassen von ihrer Bearbeitung her folgende drei Themenkreise:

I. Prüfung auf Normalverteilung

bei folgenden Hypothesen:

1. Die Systeme sind ihrem absoluten Umfang ebenso wie die Bereiche ihrem relativen Umfang nach normalverteilt.
2. Die Systemenden sind normalverteilt.

II. Prüfung auf Unabhängigkeit

bei folgenden Hypothesen:

3. Jeder relative Bereichsumfang ist unabhängig vom absoluten Systemumfang.
4. Systemanfang und -ende sind unabhängig voneinander.
5. Systemanfang und -ende sind unabhängig von den relativen Bereichsgrößen.

III. Der Streuungsvergleich

bei folgender Hypothese:

6. Die intraindividuelle Streuung ist überaus signifikant kleiner als die interindividuelle Streuung.

Für das Verständnis der Hypothesen und der Ergebnisse dieser Arbeit ist eine kurze Erläuterung der verwendeten Begriffe angebracht:

Bei den Versuchen zur Bezugssystemproblematik wurde den Vpn eine kontinuierlich abgestufte Skala von Gegebenheiten zur Beurteilung vorgelegt. Für diese Beurteilungen wurden ihnen zusätzlich die Kategorien an die Hand gegeben, nach denen die Gegebenheiten beurteilt werden sollten.
Das System oder Bezugssystem einer Vp erstreckt sich vom größten bis zum kleinsten Urteil (bei Größenbeurteilungen). Mit absolutem Systemumfang bezeichnet man die Anzahl der Versuchsgegebenheiten, die zu dem

System einer Vp gehören. Innerhalb dieser Bezugssysteme unterscheidet man einzelne Bereiche, die von den vorgegebenen Kategorien eingenommen werden. Die Anzahl der Gegebenheiten, die bei der Beurteilung mit ein und derselben Kategorie bedacht wurden, nennt man absoluten Bereichsumfang. Setzt man den absoluten Bereichsumfang in Beziehung zum absoluten Systemumfang, so erhält man den relativen Bereichsumfang. Mit Systemenden bezeichnet man die Anzahl der Versuchsgegebenheiten, die nicht mit einem kategorialen Urteil versehen wurden, die nicht mehr zum eigentlichen System der Vp gehören. Der Systemanfang ist dabei die Anzahl der Gegebenheiten, die zu klein für das System sind, das Systemende die Anzahl der Gegebenheiten, die zu groß für das System sind (bei Größenbeurteilungen).
Unter intraindividueller Streuung versteht man die Beurteilungsunterschiede einer Vp, die entstehen können, wenn diese Vp eine Gegebenheit mehrmals beurteilt; unter interindividueller Streuung versteht man die Beurteilungsunterschiede, die entstehen, wenn mehrere Vpn eine Gegebenheit beurteilen.

Die Überlegungen und Nachprüfungen in dieser Arbeit führten zu folgenden Ergebnissen:

I. Bei der Prüfung auf Normalverteilung

Die Hypothesen 1 und 2 besagen, daß die relativen Bereichsumfänge, der absolute Systemumfang und die Systemenden normalverteilt sein sollen. Da die zugrundeliegenden Versuche mit 15 bis 25 Vpn durchgeführt worden sind, konnte eine statistische Prüfung mit Aussagekraft bezüglich der Hypothesen nicht durchgeführt werden. Die Beantwortung dieser Hypothesen steht deshalb noch aus. Vorgenommen wurde lediglich eine Art Trendanalyse, die keinen Anspruch auf eine statistische Aussage erheben darf, sondern nur ein Wegweiser für spätere

Untersuchungen und Versuchsplanungen zu diesem Problemkreis darstellen soll. Bei dieser Trendanalyse stellten sich folgende Punkte heraus:

(1) Die Verteilung der Systemenden und des Systemumfanges werden durch das Nicht-Verwenden echter Systembegrenzungen (die Systeme der Vpn erstreckten sich über den gesamten dargebotenen Variationsbereich, es gab keine zu große oder zu kleine Gegebenheit) empfindlich beeinträchtigt.

(2) Die Verteilung der relativen Bereichsumfänge wird nicht von dem Vorhandensein echter Systembegrenzungen beeinträchtigt, gleichgültig ob keine, nur eine oder beide Begrenzungen vorhanden sind.

Untersuchungen, die sich mit der Prüfung auf Normalverteilung befassen wollen, sollten unbedingt mit wenigstens 1oo Vpn durchgeführt werden. In Ausnahmefällen reichen auch 5o Vpn, nur machen sich dann Unregelmäßigkeiten bei der Verwendung der Systemenden besonders stark bemerkbar (das Nicht-Vorhandensein einer Systembegrenzung bedeutet eine Häufung in der Extremklasse). Allgemein gilt:
Je größer die Anzahl der Vpn, desto weniger empfindlich werden die Versuchsergebnisse gegenüber Systembegrenzungsunzulänglichkeiten.

II. Bei der Prüfung auf Unabhängigkeit

Die Hypothesen 3, 4 und 5 besagen, daß der relative Bereichsumfang unabhängig vom Systemumfang sei, der Systemanfang vom -ende und die Systemenden vom Bereichsumfang. Postuliert wird hier die Nullkorrelation zwischen jeweils zwei Größen. Die statistische Verfahrensweise bei der Korrelationsrechnung geht von der Hypothese aus : es besteht ein Zusammenhang zwischen zwei

Größen. Da hier die Umkehrung dieser Hypothese geprüft werden soll (es besteht kein Zusammenhang), können die statistischen Entscheidungskriterien nicht angewendet werden. Aus diesem Grunde wurden Entscheidungshilfen konstruiert, die aber nicht mit Sicherheit zu der Aussage 'kein Zusammenhang' führen können. Bekannt sind die Wahrscheinlichkeitswerte der Nullkorrelation durch die zugehörige Normalverteilung (Mittelwert gleich Null), erforderlich wären die Wahrscheinlichkeitswerte für die Korrelation $\pm$ 1. Die Normalverteilung über dem Mittelwert $\pm$ 1 ist jedoch leider nicht bekannt.
Die Normalverteilung über die Nullkorrelation gibt an, wann ein statistisch gesicherter Zusammenhang besteht; außerdem kann man über die σ-Werte den Bereich ermitteln, in dem 68% der Nullkorrelationen liegen. Auf Grund der errechneten Korrelationskoeffizienten und der zugehörigen Wahrscheinlichkeitswerte aus der Normalverteilung über der Nullkorrelation lassen sich folgende Aussagen machen:

(1) Die Mittelwerte der Korrelationskoeffizienten zeigen deutlich eine Tendenz zur Nullkorrelation, also zu Unabhängigkeit.

(2) Nur relativ wenige Koeffizienten zeigen gesicherten Zusammenhang an, die meisten Koeffizienten lagen innerhalb der Streugrenzen $\pm$ 1 σ (Mittelwert = o).

(3) Es trat ein Konstanteneffekt durch die nicht verwendeten Systemenden auf (hierbei ist der Wert für den Systemanfang oder das -ende bei vielen Vpn gleich Null, wodurch eine scheinbare Unabhängigkeit erreicht wird).

Da die statistischen Prüfverfahren und ihre Entscheidungskriterien von der Frage 'Zusammenhang oder nicht' ausgehen, müssen bei späteren Versuchsplanungen die darausfolgenden Abhängigkeitsverhältnisse beachtet werden. Die Wahrscheinlichkeitswerte, die zu der Annahme der Hypothese 'Zusammenhang' führen, sind so angelegt, daß dies bei wenigen Vpn immer schwieriger wird. Geht man nun von der Umkehrung dieser Hypothese aus, wie es bei der Fra-

ge nach Unabhängigkeit der Fall ist, so ist diese bei wenigen Vpn kaum mehr zu widerlegen. Man muß deshalb die Versuche mit so vielen Vpn durchführen, daß allein vom Wahrscheinlichkeitsrechnerischen Standpunkt aus beide Möglichkeiten, Zusammenhang oder Unabhängigkeit, die gleich Chance haben. Dies ist erreicht bei Versuchen mit ca. 3o bis 4o Vpn.

Bei allen bisher behandelten Hypothesen (1 bis 5), ob sie nun Normalverteilung oder Unabhängigkeit besagten, tritt das Problem der Systembegrenzungen auf. Außer bei der Prüfung der relativen Bereichsumfänge auf Normalverteilung zeigen die Ergebnisse immer wieder eine starke Abhängigkeit vom Vorhandensein echter Systembegrenzungen.

Ist ein Versuch so angelegt, daß die Variationsbreite der Gegebenheiten nicht ausreicht und viele Vpn ein Systemende gar nicht oder nur einmal verwenden, so kann dieses Systemende nicht normalverteilt sein, da eine Häufung in der Extremklasse vorliegt. Außerdem kann dieses Systemende keine echte Unabhängigkeit von einer anderen Größe erzielen, da bei der Korrelationsrechnung die Korrelation mit einer Konstanten, was das Systemende dann ja fast darstellt, stets gleich Null ist. Die Unabhängigkeit, die man dadurch erzielt, ist eine Scheinunabhängigkeit, die in dem angewandten Verfahren begründet liegt und nicht in Eigenschaften der Bezugssysteme.

Interessant war die Untersuchung der Frage, warum bei den vorliegenden Versuchen einzelne Systemenden von vielen Vpn nicht verwendet wurden. In vielen Fällen war der Grund dafür in unterschiedlichen Beurteilungsaspekten zu finden:

Bei den Bleistiftversuchen war der Großbereich durch den Normbleistift festgelegt. Im Bereich der kleinen und kleinsten Bleistifte konnten bei der Beurteilung zwei Gesichtspunkte zur Wirkung kommen: einmal der Bleistift in seiner Funktion als Schreibwerkzeug, zum anderen der Blei-

stift in seiner äußeren Erscheinungsform als Holzstümmelchen. Beurteilt man nach dem Gesichtspunkt 'Schreibwerkzeug', so ist ein zu kleiner Stift nicht mehr zu gebrauchen und muß abgelehnt werden. Beurteilt man aber nach der äußeren Form, so ist auch der kleinste Stummel noch ein Stück Holz mit einer Bleimine und kann nicht abgelehnt werden.

Diese Beurteilungsunterschiede haben einen direkten Einfluß auf das Systemende. Wird nach der Funktion geurteilt, so ist ein Ende vorhanden. Urteilt man aber nach dem Aussehen, so akzeptiert man alle Stifte und es ist kein Systemende vorhanden. Da im Großbereich nicht die Möglichkeit der Systemenden bestand, weil hier die Variationsbreite durch den Normbleistift beendet wurde, ergibt sich auch ein Zusammenhang mit dem Gesamtumfang, der sich aus der Variationsbreite minus dem Systemende ergibt. Ist ein Systemende da, ist der Umfang kleiner als wenn kein Systemende da ist, weil er dann gleich der gesamten Variationsbreite ist. Aus diesem Grunde schälte sich zusätzlich ein positiver Zusammenhang zwischen der Größe des Umfanges und der Größe des Kleinbereiches heraus. Ist der Kleinbereich klein, so ist auch der Umfang klein und umgekehrt.

Bei späteren Versuchen bemühte man sich, den Beurteilungsgesichtspunkt durch die Versuchsanweisung festzulegen, damit nur ein Aspekt zur Wirkung kommen könne. Bei den Versuchen mit Armbanduhren und Taschentüchern weisen die Ergebnisse auch nicht wieder auf Schwächen bei den Systembegrenzungen hin. Bei den Versuchen mit den Christbaumkugeln legen die Ergebnisse jedoch wieder den Schluß nahe, daß bei der Beurteilung unterschiedliche Aspekte zur Wirkung gekommen sind. Die Versuchsanweisung will die Beurteilung in einer Richtung festlegen, indem sie die Vp auffordert, die Größe der Kugeln in Hinblick auf den allgemein üblichen Weihnachtsbaum zu beurteilen. Dieser Hinweis auf das allgemein Übliche kann aber die Vp in gewisse Konflikte bringen. Sie besitzt ein persönliches Bezugssystem für die Baum- und damit die Kugelgröße,das nicht unbedingt mit dem allgemeinen Stan-

dard übereinstimmen muß. Durch diesen ausdrücklichen Verweis auf die Norm kann einmal die Spontanität der Beurteilung beeinträchtigt werden, zum anderen fällt es den Vpn schwer, im Großbereich Kugeln abzulehnen, weil man sich immer vorstellen kann, daß es eine so große Kugel an irgendeinem Baum geben könne, zumal die vorgegebenen Kugeln eine gewisse Extremität der Größe nicht erreichen.

Diese Überlegungen zeigen, daß man bei der Planung folgender Versuche darauf achten muß, echte Systembegrenzungen zu erhalten. Dies sollte man in entsprechenden Vorversuchen abklären, da oft Beurteilungsaspekte miteinfließen können, an die man bei der Planung gar nicht gedacht hat (siehe Christbaumkugeln) und die sich bei Schwächen in den Begrenzungen auswirken können.

III. Beim Streuungsvergleich

Bei dem Vergleich der interindividuellen Streuung mit der intraindividuellen Streuung spielte das Problem der Systembegrenzungen eine nebengeordnete Rolle. Unbedingt notwendig für diesen Vergleich sind lediglich Versuchswiederholungen, die bislang nicht sehr oft durchgeführt worden sind.

Für die Ermittlung der Streuwerte gibt es keine allgemein üblichen statistischen Prüfverfahren. In dieser Arbeit wurde deshalb einmal nach der BRÄUERschen Methode vorgegangen, die direkt von den gegebenen Urteilen ausgeht und Stabilitäts- und Labilitätspunkte (zusammen gleich 100%) errechnet, und zum anderen eine Methode von Mittelwert und Streuung erarbeitet, die nicht direkt auf den gegebenen Urteilen basiert, sondern auf deren Mittelwerten. Bei der BRÄUERschen Methode können die Labilitätspunkte als Maß für die Streuung angesehen werden, bei der anderen Methode die Abweichungen der Urteile vom errechneten Mittelwert.

Beide Methoden führen zu dem gleichen Ergebnis, nämlich daß die intraindividuelle Streuung signifikant kleiner ist als die interindividuelle Streuung. Im einzelnen erbrachten beide Methoden folgende Ergebnisse:

Methode von BRÄUER:

(1) Versuche mit drei Kategorien weisen keine intraindividuelle Streuung auf.

(2) Bei den Versuchen mit fünf Kategorien war die intraindividuelle Streuung signifikant kleiner als die interindividuelle.

(3) Es stellte sich eine Abhängigkeit der Labilitätspunkte von der Anzahl der Kategorien dahingehend heraus, daß bei größerer Kategorienanzahl auch die Anzahl der Labilitätspunkte zunahm (diese Abhängigkeit veranlaßte die Ausarbeitung der zweiten Methode).

Methode von Mittelwert und Streuung:

(1) Die intraindividuelle Streuung war stets kleiner als die interindividuelle.

(2) Der Bereich, in dem die intraindividuellen Streuungen liegen, ist deutlich abgehoben von dem Bereich, in dem die interindividuellen Streuwerte liegen.

(3) Die Verteilung der Ergebnisse stellte sich als sehr homogen heraus, es gab keine Unterschiede entsprechend der im einzelnen Versuch verwendeten Kategorienanzahl.

Anschließend wurde noch ein Vergleich der beiden Methoden durchgeführt. Die Tendenz der Ergebnisse erwies sich als gleich, lediglich die Größenordnung der Streuwerte und die Lage der Bereiche von intraindividueller zu interindividueller Streuung waren unterschiedlich. Ein direkter Vergleich an einer ausgewählten Versuchsreihe in Bezug auf altersmäßige Schwankungen ergab, daß inhaltlich keine Unterschiede bestehen.

Die Abhängigkeit der Streuwerte bei der BRÄUERschen Methode von der Anzahl der verwendeten Kategorien ist ein Nachteil dieser Methode. Da sie aber direkt von den Urteilen ohne

weitere Verrechnung ausgeht und somit ohne Informationsverlust arbeitet, sollte man diese Methode weiter ausbauen und versuchen, die Abhängigkeit von der Anzahl der Kategorien in den Griff zu bekommen.

Eine interessante Tatsache ergab sich bei der BRÄUERschen Methode bei den Versuchen mit drei Kategorien. Für die Ermittlung der Labilitätspunkte ergab sich das Problem, wie die Systemenden zu behandeln seien. Sollten sie als Kategorie angesehen werden und wie diese mit in die Berechnung eingehen, oder sollte man sie ganz herauslassen? Es wurden beide Möglichkeiten durchgerechnet und man erhielt ein interessantes Ergebnis:

Bei den Drei-Kategorien-Versuchen hatte sich ohne die Systemenden keine intraindividuelle Streuung ergeben, bei den Fünf-Kategorien-Versuchen jedoch immer.Nahm man nun die Systemenden bei der Berechnung wie Kategorien mit hinein, so erhielt man dadurch eigentlich Fünf-Kategorien-Versuche. Es ergab sich aber auch dann keine intraindividuelle Streuung. Die Behandlung der Systemenden wie Kategorien ist somit nicht statthaft, da sie anscheinend etwas anderes darstellen als diese. Was das Wesen der Systemenden und in Abhebung davon das der Kategorien ist, wäre eine interessante Forschungsaufgabe. Im Rahmen dieser Arbeit konnte leider nicht weiter darauf eingegangen werden.

Literaturverzeichnis

1) HEHLMANN, W. : Wörterbuch der Psychologie, 1965

2) METZGER, W. : Psychologie, 1954

3) WITTE, W. : Struktur, Dynamik und Genese von Bezugssystemen, Psychologische Beiträge Bd. IV, Heft 2, 196o, 218 -252

4) HRUSCHKA, E. : Experimentelle Untersuchung zur Struktur von eindimensionalen Bezugssystemen, Diss. Tübingen 1959

5) BRÄUER, K. : Die Entwicklung von Bezugssystemen von der späten Kindheit an über Vorpubertät und Pubertät zum Erwachsenenalter, Diss. Münster 1971

6) BUDDE, H.-G. : Experimentelle Untersuchung über Partialsysteme musikalischer Töne, Zulassungsarb. z. Dipl.-Vorprüfung für Studierende d. Psychologie: Münster 1971

7) WINKELMANN, R, : Verbale Kategorisierung bei Kindern, Diss. Tübingen 1966

8) ALTRICHTER, E. : Experimentelle Untersuchung des in arithmetischer Abstufung gegebenen Bezugssystems Grauleiter, Zulassungsarb. z. Dipl.-Vorprüfung für Studierende d. Psychologie: Tübingen 1956

9) STEINER, W. : Experimentelle Untersuchung zur Struktur des in geometrischer Abstufung gegebenen Bezugssystems Grauleiter, Zulassungsarb. z. Dipl.-Vorprüfung für Studierende d. Psychologie: Tübingen 1955

1o) GUILFORD, J.P. : Fundamental Statistics in Psychology and Education, 1965

11) MITTENECKER, E. : Planung und statistische Auswertung von Experimenten, 1966

LEBENSLAUF

Am 28. 3. 1947 wurde ich als Tochter des Diplom-Kaufmannes Dr. jur. Erich Zahn und seiner Gattin Gertrud, geb. Folle, in Herford geboren.

Ostern 1953 wurde ich in Gelsenkirchen in die dortige evangelische Volksschule eingeschult, die ich 4 Jahre besuchte. Ostern 1957 wechselte ich zum Ricarda-Huch-Gymnasium, einem neusprachlichen Mädchengymnasium über, das ich 9 Jahre lang besuchte. Am 26. 2. 1966 beendete ich meine Schulzeit nach bestandenem Abitur.

Mein Studium begann im Sommersemester 1966 an der Maximilians-Universität München mit dem Studienfach Volkswirtschaft. Nach einem Semester wechselte ich das Studienfach und die Universität. Im Wintersemester 66/67 begann ich an der Wilhelms-Universität Münster mein Psychologie-Studium. Nach dem 5. Semester, am 26. 2. 1969, legte ich das Vordiplom, nach dem 8. Semester, am 29.1o.197o, mein Hauptdiplom ab.

Während meines Studiums arbeitete ich am Psychologischen Institut als Tutor für Statistik und als studentische Hilfskraft. Nach bestandenem Diplom setzte ich diese Tätigkeit als wissenschaftliche Hilfskraft fort, mein Aufgabengebiet war die Programmierung Psychologischer Probleme, in dem ich auch 2 Semester lang Lehrveranstaltungen abhielt. Meine Tätigkeit am Institut endete im November 1972.

Im August 1971 regte Herr Prof. Witte mich zu der vorliegenden Dissertation an, an der ich bis zum August 1973 arbeitete. Während dieser Zeit erhielt ich vom Nov. 1972 bis einschließlich Okt. 1973 ein Stipendium im Rahmen des Graduierten-Förderungs-Gesetzes.

Am 11. August 1972 heiratete ich den cand. med. Rüdiger Forner. Seit Okt. 1973 arbeite ich als freie Mitarbeiterin, eine Tätigkeit, die ich auch schon nach meinem Examen 1 Jahr lang neben der Tätigkeit am Psychologischen Institut durchführte.

Corinna Forner